Modelling microscopic clusters of sulphuric acid and water relevant to atmospheric nucleation

Jake Lewis Stinson

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Physics and Astronomy

University College London

9th January 2015

I, Jake Lewis Stinson confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

To my family, friends and colleagues for their unwavering support and exceptional patience.

Abstract

Classical nucleation theory has been a useful tool for predicting the phenomena of nucleation for the past seventy years. However the model has several limitations, which in some examples give rise to predicted rates that are several orders of magnitude in error. One such example is that of sulphuric acid and water nucleation which has long been framed as an important source of cloud condensation nuclei and therefore has implications for the climate, both locally and globally. In addition stratospheric aerosol injection of molecules containing sulphur, including sulphuric acid, are of interest as a potential geoengineering technique. The focus for this study is to improve upon our understanding of water and sulphuric acid nucleation.

The initial phase of the project concerned performing quantum chemistry calculations which go beyond the macroscopic description employed by classical nucleation theory. Kohn-Sham density functional theory has been successfully employed in the fields of condensed matter, material physics and chemistry. However one of the assumptions of the theory is the classical treatment of the nuclei. The path integral molecular dynamics technique is used here to test this assumption on small clusters of sulphuric acid and water. We find that the introduction of zero point motion has a small effect on the equilibrium properties of certain configurations in line with expected behaviour. An interesting structure is found which serves to emphasise the importance of liquid like behaviour in this cluster at room temperature.

The first study demonstrated the computational expense of treating systems at the microscopic scale using quantum chemistry approaches. The second phase of the research focused upon finding a suitable classical potential to employ within a molecular dynamics scheme, which would drastically reduce the computational expensive of performing simulations. This potential would be required to retain the ability for protons to transfer between selected species. The empirical valence bond method was chosen for its straightforward implementation and its similarity to traditional classical schemes. However some modifications were required to implement the scheme. Two algorithms were designed to identify species within the system and treat them in a fashion suitable for use in the empirical valence bond method.

In addition the empirical valence bond method also needed to be parametrised for the sulphuric acid and water system. This was achieved by using the particle swarm optimisation technique, which performed force matching parametrisation using the Kohn-Sham density functional theory work from the previous phase of the project. The model was fully programmed in FORTRAN 90/95 and was incorporated into DL-POLY version 4.03. It is tested against the density functional theory data to which it is parametrised to check that the main features of the quantum chemistry are retained within the empirical valence bond technique.

A puzzling issue appeared in preliminary molecular dynamics simulations performed with DL-POLY 4.03. The issue arises from a constraint imposed to fix the centre of mass. The solution to the modified Langevin equation introduced by this constraint is derived. The results are then compared to the puzzling DL-POLY simulations and found to be consistent. The constraint is then removed for all further simulations.

The developed empirical valence bond potential was used to perform simulations of small clusters of sulphuric acid and water. We test the level of hydration required to ionise the system and find it to be in line with literature values. A thermodynamic integration scheme that was suitable for this system was derived. Preliminary simulations were performed using the model to compute free energies for use with classical nucleation theory in order to calculate nucleation rates.

Acknowledgements

I thank my supervisor, Prof. Ian Ford, for his unwavering guidance and exceptional patients. I thank Dr. Shawn Kathmann and Prof. Angelos Michaelides for support and useful discussion concerning the direction of the project. I am grateful for discussions I had with Prof. Hanna Vehkamäki, Dr. Theo Kurtén, Prof. Gregory Voth, Dr. Gregory Schenter, Dr. Dorothy Duffy, and Dr. Peter Sushko concerning several aspects of the project.

My friends and family have provided vital support throughout the project for which I thank them. I draw special attention to my parents for their support, both financial and emotional. My friends have provided a supportive atmosphere for my studies and I acknowledge David Buckley, Michael Davis and Matthew Darby for their time and kindness.

This project was supported by the IMPACT scheme at UCL and by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences. I am grateful for use of the UCL Legion High Performance Computing Facility and the resources of the National Energy Research Scientific Computing Center (NERSC), which is supported by the U.S. Department of Energy, Office of Science under Contract No. DE-AC02-05CH11231.

Contents

List of Figures

5 Path integral molecular dynamics

6 Empirical Valence Bond model

7 Parametrisation of the EVB model

8 Free energy calculation

Appendices

A Canonical temperature distribution

List of Tables

Publications

The Ph.D. project produced the publications listed here.

J. Stinson, S. Kathmann and I. Ford "Dynamical consequences of a constraint on the Langevin thermostat in molecular cluster simulation" *Molecular Physics*, pp. 1-4 (2013), doi:10.1080/00268976.2014.917732

J. Stinson, S. Kathmann and I. Ford "Empirical valence bonds: A reactive classical potential for sulphuric acid and water", *AIP Conference Proceedings*, **1527**(1), pp. 266-269 (2013), doi:10.1063/1.4803255

J. Stinson, S. Kathmann and I. Ford "Investigating the significance of zero-point motion in small molecular clusters of sulphuric acid and water" *The Journal of Chemical Physics*, **140**(2), 024306 (2014), doi:10.1063/1.4860973

Chapter 1

Introduction

1.1 Overview of thesis

This project started with the clear aim to improve our understanding of the binary nucleation of gaseous sulphuric acid and water which is relevant to atmospheric chemistry and global climate. Over the course of this undertaking there have been several studies and reports. The first chapter provides an introduction to the relevance of sulphuric acid and water nucleation and classical nucleation theory which provides a framework for modelling nucleation events. Chapter **2** continues with a literature review of the experimental and theoretical efforts to study sulphuric acid and water. The theoretical review starts with quantum chemistry studies of clusters of sulphuric acid and water which offer high accuracy at the cost of computational demand. Classical approaches are then reviewed with reactive methods being given extra attention as proton transfer is an important mechanism in clusters of sulphuric acid and water. The third chapter gives a background on running molecular dynamics simulations and several useful theoretical tools for performing analysis. An issue with Langevin thermostats is investigated in this chapter with this work being published in reference [**1**]. Chapters **4** and **5** review the methods of electronic density functional theory and path integral molecular dynamics. These techniques are then used to study small clusters of sulphuric acid and water using these techniques, the results are also discussed in reference [**2**]. In the sixth chapter the empirical valence bond method is introduced and is adapted for use with the sulphuric acid and water system. The model is then parametrised in the following chapter using the particle swarm optimisation technique. The particle swarm optimisation is first used on a test case and then used to fit the parameters of the empirical valence bond method as introduced in the previous section. Chapter **8** presents the thermodynamic integration method and some preliminary work is performed using the

parametrised model to calculate free energies of small clusters of sulphuric acid and water using this technique with the empirical valence bond potential developed in the previous two chapters. The thesis is concluded in Chapter **9** with discussion of the studies performed here and final thoughts on the future work.

1.2 Atmospheric aerosols and geoengineering

Atmospheric aerosols are particles in a solid or liquid like state which are suspended in the atmosphere. They vary in diameter; from a few nanometres to hundreds of micrometres. Aerosols are a major factor in the earth's balance of solar radiation due to their light scattering properties. They affect atmospheric chemistry, an example of this is the depletion of the stratospheric ozone layer [**3**]. Some aerosols act as cloud condensation nuclei, CCN's [**4**] which are a key component for water to transition from the gas to the liquid phase in atmosphere.

Global warming is a topic which is increasingly becoming important for many countries. The recent Cancún agreement in 2010 set a target of a maximum warming of 2°C over the pre-industrial global mean temperature. A reduction in emissions of greenhouse gases is the most straightforward method for achieving this, however, alternatives to this plan are necessary if a reduction in emissions fails to be realised. One such contingency being considered is that of geoengineering which includes techniques such as carbon capture and injecting aerosols into the stratosphere.

Volcanic eruptions are a striking example of how aerosols can directly affect the environment. Such eruptions expel gases into the atmosphere which can nucleate into a variety of aerosols and can affect the global climate, of which sulphur containing species are of particular interest [**6**]. Geoengineering methods inspired by the volcano example offer a possible short term alternative to emission cuts for limiting temperature rises. An extensive review of a variety of geoengineering options including stratospheric aerosols ejection has been produced by the Royal Society [**5**]. Figure **1.1** displays the *'provisional overall evaluation'* from the review, which shows that stratospheric aerosols are a promising geoengineering technique. The focus of stratospheric aerosol methods has been on sulphur containing species due to evidence of their effectiveness coming from volcanic eruptions and that they can be introduced to the atmosphere as a gas which is more practical for achieving an even regional distribution as required.

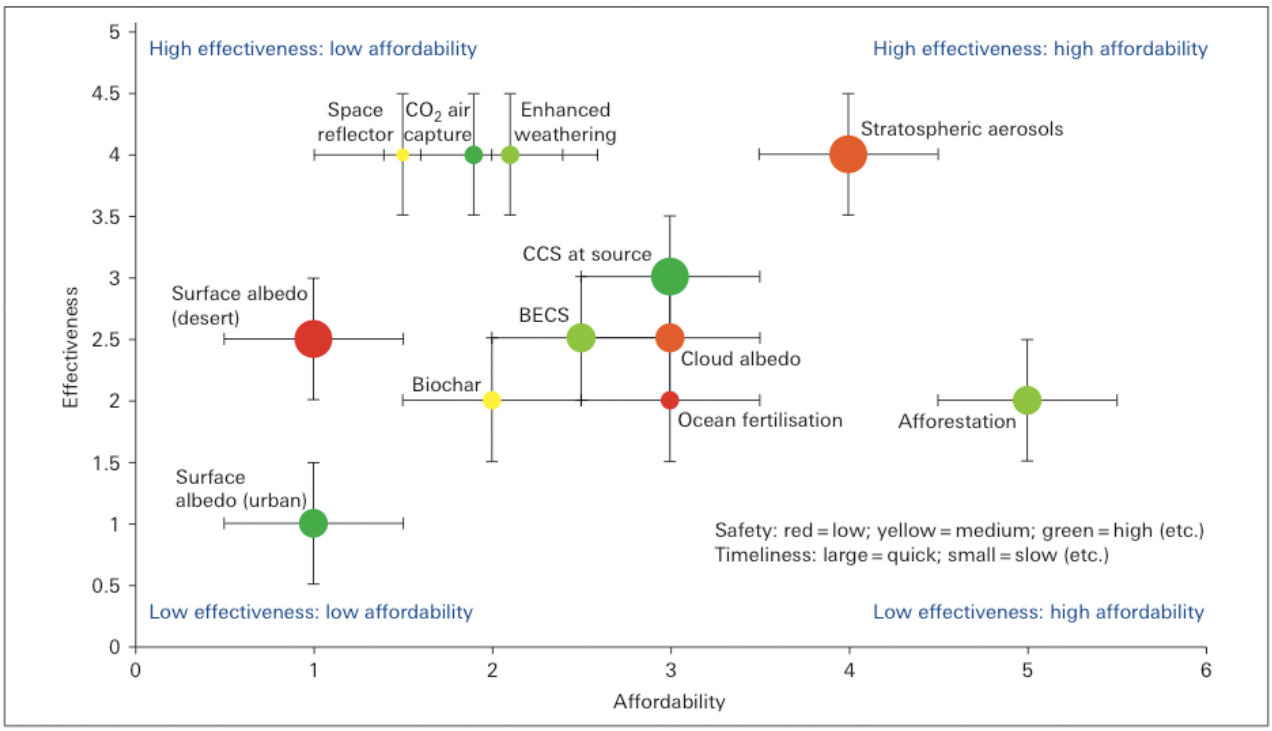


Figure 1.1: Figure 5.1 in Geoengineering the climate: Science, governance and uncertainty **[5]**. The size of points indicate their implementation timescale; where a large point indicates a swift effect and a small point represents a slow effect. The colour of each point represents their safety with green points being the safest and red points being the least safe. The figure is described as *'provisional overall evaluation'*, it is clear that stratospheric aerosols are considered to be effective, affordable and to be rapidly implementable, however their safety is a concern.

1.3 Sulphuric acid

Sulphuric acid has the molecular formula $[H_2SO_4]$, it is a strong acid as it rapidly reacts with water as $[H_2SO_4] + [H_2O] \rightarrow [HSO_4]^- + [H_3O]^+$ to form a bisulphate ion $[HSO_4]^-$ and a hydronium ion $[H_3O]^+$. The structure of an isolated sulphuric acid molecule is displayed in Figure **1.2**. Sulphuric acid is thought to be the most prominent atmospheric nucleating agent **[7, 8]** on account of its low saturated vapour pressure, relatively high atmospheric concentration and its affinity to water. The saturated vapour pressure refers to the pressure of the vapour in thermodynamic equilibrium with its condensed phase within a closed system. Recent developments (see 2.2.1) have led to questions surrounding the ability of sulphuric acid alone to explain observed aerosol formation in the atmosphere **[9]**. Ternary species such as ammonia **[10]** and organic molecules such as dimethylamine **[11]** have been suggested as agents which can enhance the nucleation rates and lead to better agreement between experiment and observation.

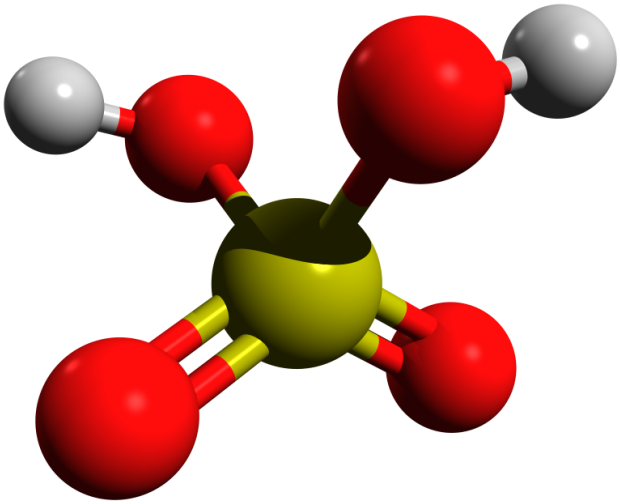


Figure 1.2: An image of a sulphuric acid molecule.

To understand sulphuric acid's role in the nucleation of water vapour it is crucial to have an accurate grasp on the free energy landscape of different sulphuric acid and water clusters. The free energy of a cluster is a thermodynamic property of the system **[12]** and is discussed further in the next section.

1.4 Free energy

A key quantity in thermodynamics is that of free energy, φ, which is equivalent to the quantity of work which a thermodynamic system can perform in a reversible isothermal process. The constraints imposed on a system (such as holding the number of particles in a system constant) result in different relevant free energies for that particular system, commonly used constraints and free energies are detailed in Table **1.1**. They are decomposed into several macroscopic quantities of a system including: internal energy, U; temperature, T; entropy, S; pressure, P; volume, V; Number of particles, N and the chemical potential, μ.

The classical picture of a phase transformation is that shown in Figure **1.3** where the free energy

Free Energy	**Definition**	**Constraints**
Helmholtz	$\mathcal{F} = U - TS$	NVT
Gibbs	$\mathcal{G} = U + PV - TS$	NPT
Grand Potential	$\Phi_G = U - TS - \mu N$	μVT

Table 1.1: The definition of several commonly used free energies.

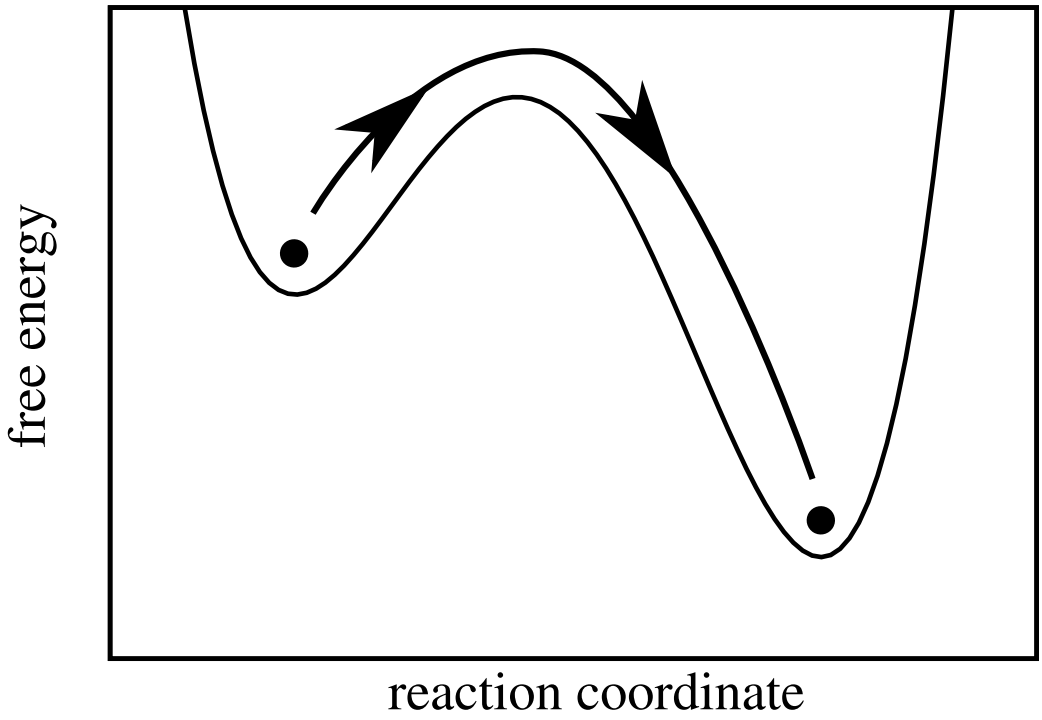


Figure 1.3: An example of a free energy landscape for an arbitrary system. To reach the global minimum it must climb over the barrier which requires a rare fluctuation. The height of the barrier with respect to the local minimum is the most significant property controlling the rate at which a barrier crossing event can occur.

of the system varies according to the state. It is clear that the global minimum in the free energy is on the right hand side, however, if a system finds itself in the localised free energy minimum on the left hand side it has to increase its free energy before it can transit to the global minimum in free energy. The properties of this barrier are crucial for studying the rate at which a system may cross. For instance, there may be an entropic cost associated with an unfolding of a protein at constant volume and temperature, in this situation the applicable free energy is that of Helmholtz. For nucleation the free energy of interest is typically composed of the positive surface term and the negative bulk term which competing effects give rise to the barrier which defines a nucleation[1]. Classical nucleation theory forms the basis for a large number of theories which predict nucleation and is presented in the next section.

[1] A barrier-less phase transition is not a nucleation process as it happens immediately.

1.5 Classical Nucleation Theory

Here classical nucleation theory is reviewed and the free energy calculations are identified as an area where it is possible to go beyond classical nucleation theory. Improving upon these free energy calculations forms the main theme of the thesis.

1.5.1 Classical Nucleation Theory and the Arrhenius equation

Nucleation as a phenomenon of phase transitions has been known for at least 300 years with Fahrenheit in 1724 finding that the freezing point of water is dependent on the freezing condition **[13]**. Further information on the early works of nucleation can be found in the reviews by Kathmann **[14]** and Vehkamäki **[13]**. Classical Nucleation Theory (CNT) was derived in the first half of the 20^{th} century by Becker and Döring **[15]** (1935) and Zeldovich **[16]** (1942). It is convenient to start a derivation of CNT with the Arrhenius equation **[17]**,

$$k(T) = A \exp\left[\frac{-E_{act}}{RT}\right] \tag{1.5.1}$$

where $k(T)$ describes the rate of a reaction in units of $[\mathrm{s}^{-1}]$, R is the ideal gas constant with units of $[\mathrm{JK}^{-1}\mathrm{mol}^{-1}]$, A is commonly referred to as the frequency factor as it can be described as the number of collisions that result in a reaction and has units of $[\mathrm{s}^{-1}]$ in a first order reaction, it may have a temperature dependence. For higher order reactions A has units of $[\mathrm{concentration}^{1-\mathrm{p}}\mathrm{s}^{-1}]$ where p is the order of the reaction. E_{act} is the activation energy which can be thought of as the minimum amount of energy required for a reaction to take place. Importantly E_{act} is equivalent to the height of free energy barrier (per mol*e*) between the reactant state and the product state of the reaction of interest. It is possible to reformulate Eq. (**1.5.1**) to relate the equilibrium populations of a cluster of i of the same molecules[2] (referred to as an i-mer with a population of $\rho_e(i)$) to that of the equilibrium monomer number density ($\rho_e(1)$) such that,

$$\rho_e(i) = \rho_e(1) \exp\left[\frac{-\Delta\varphi_i}{k_B T}\right] \tag{1.5.2}$$

[2]This is analogous to the Arrhenius equation as the reaction can be seen as the formation of the liquid cluster out of the monomer vapour.

where $\Delta\varphi_i$ is the free energy difference between a monomer and an i-mer. k_B is the Boltzmann constant with units of $[\mathrm{JK}^{-1}]$. In the single component case $\Delta\varphi$ is expressed in CNT (where the system is that of a vapour condensing into the liquid phase) as,

$$\Delta\varphi = i\Delta\mu + \sigma A(i) \tag{1.5.3}$$

where $\Delta\mu$ describes the difference in chemical potential of a liquid and a vapour in equilibrium where both quantities are at vapour pressure. i denotes the number of molecules within the system. σ and $A(i)$ represent the surface tension and surface area of the i-mer respectively. Eq. (**1.5.3**) is often simplified by the capillary approximation where the surface tension is assumed to be that of a planar surface. Eq. (**1.5.3**) is often expressed in terms of measurable quantities as,

$$\Delta\varphi = \sigma\left(36\pi i^2 v_l^2\right)^{1/3} - ik_BT\ln\left(\mathcal{S}\right) \tag{1.5.4}$$

where v_l is the molecular volume and $\mathcal{S}$ is the saturation ratio of the i-mer and is defined by $\mathcal{S} = P_v/P_{vs}$ where P_v and P_{vs} are the vapour pressure and the saturated vapour pressure respectively. The saturation vapour pressure describes the pressure of a vapour which is at equilibrium with a liquid within a closed system. Figure **1.4** shows $\Delta\varphi$ as a function of i according to CNT. The most important feature of Figure **1.4** is the apex of the barrier which is known as the critical cluster and its properties are closely related to the nucleation rate.

Eq. (**1.5.2**) is used to determine the population of i-mer clusters in equilibrium, however, we are really interested in a nucleation rate to describe the number of i-mers produced from a supersaturated vapour within a standardised unit of time. To determine this it is important to also consider the collision rate of molecules onto an i-mer, $\beta(i)$, this is usually simplified to considering only that of monomer collisions, $\beta_1(i)$, with the i-mer. The nucleation rate, J, is then proportional to,

$$J \propto \beta_1(i^*)\rho_e(1)\exp\left[\frac{-\Delta\varphi_{i^*}}{k_BT}\right] \tag{1.5.5}$$

where i^* denotes the critical cluster. The critical cluster plays this important role because that any further growth reduces the cluster's free energy and that after this growth it becomes favourable for the cluster to grow rather than decay. The dependence of the CNT nucleation rate on the free energy

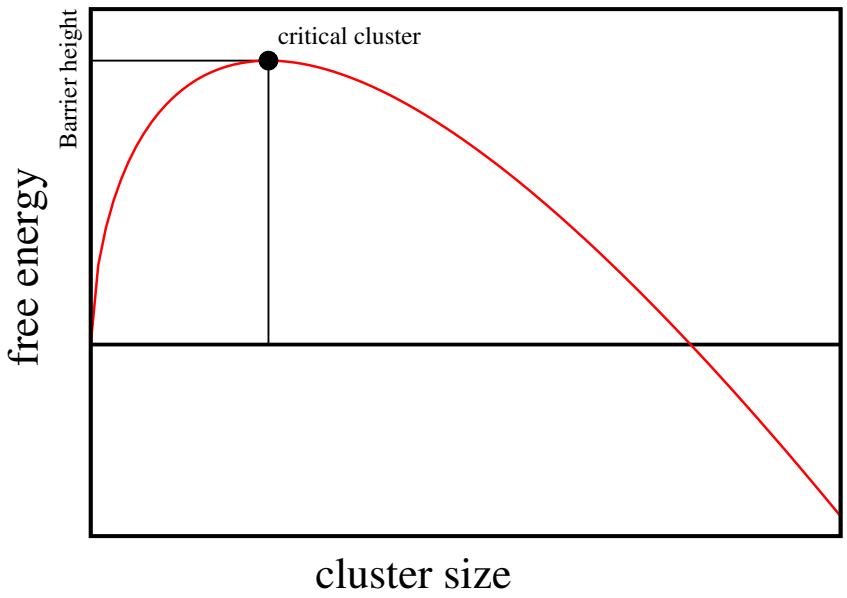


Figure 1.4: The graph shows the change in free energy against cluster size according to Eq. (**1.5.4**). The critical cluster is marked out and represents the size at which the likelihood of growth and decay become equal. Once the cluster has overcome this barrier ($N > N_{critical}$) it will grow until it has exhausted the number of monomers (and other small clusters) available to it.

barrier $\Delta\varphi_{i^*}$ is thought to be problematic, attempts to improve upon the CNT method for calculating this are discussed in Section **1.5.4**.

1.5.2 The kinetic interpretation

It is appealing to study the phenomenon of nucleation as a kinetic process where particles of the parent phase continuously collide and stick together to form clusters of the condensed phase which can then either grow or decay. These rates of growth for the cluster are known as the forward or condensation rate ($f(n)$ where n indicates the number of particles within a cluster) and the rate of decay is known as the backwards or evaporation rate ($b(n)$ where n indicates the number of particles within a cluster). The process can then be seen as study of the population of differently sized clusters and how these populations change over time. The Becker-Döring equations **[15]** are a set of coupled equations that describe this kinetic interpretation with a number of assumptions;

- Growth and decay can only change the number of particles within a cluster by one. This excludes the possibility of evaporation and condensation events involving dimers and larger clusters.

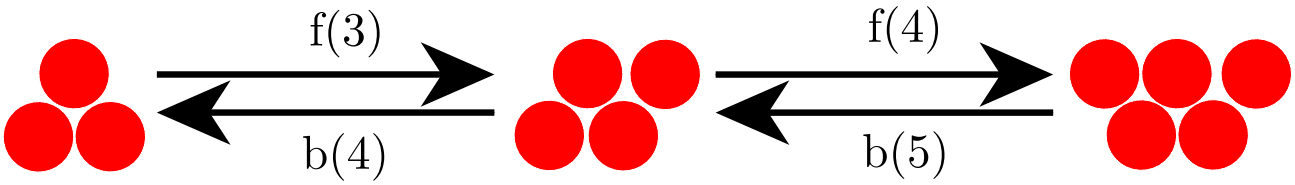


Figure 1.5: Image showing the trimer, tetramer and pentamer clusters. The tetramer cluster has four possible mechanisms for changing its population. These are specifically evaporation and condensation to and from adjacently sized clusters under the assumption that the only changes involving the monomer can occur.

- The sticking ratio is the probably that a monomer with be absorbed by the cluster upon collision and is assumed to be unity for all clusters.

- The process is assumed to be a Markov process where each cluster has no memory of its previous evaporation and condensation events.

The coupled equations then take the form of,

$$\frac{\partial \rho(n)}{\partial t} = f(n-1)\rho(n-1) - b(n)\rho(n) + b(n+1)\rho(n+1) - f(n)\rho(n) \tag{1.5.6}$$

here $\rho(n)$ is the density of n-clusters[3] within the system. The net rate by which a n-cluster becomes a $(n+1)$-cluster is then defined by,

$$\mathcal{J}_n = f(n)\rho(n) - b(n+1)\rho(n+1) \tag{1.5.7}$$

where $\mathcal{J}_n$ is known as the nucleation current. It then remains to evaluate b and f for all n-clusters under consideration.

1.5.2.1 Condensation rate

The condensation rate, $f(n)$ can be found by considering the kinetic rate of collisions between the n-cluster and monomers within the system,

$$f(n) = v_{in} A(n) \tag{1.5.8}$$

[3]Here we switch notation from i-mer to n-cluster for greater clarity in this section.

where $A(n)$ is the surface area of the cluster and v_{in} is the monomer flux and is known as the impingement rate. v_{in} can be found via gas kinetic arguments [**18**],

$$v_{in} = \frac{P_{vs}}{\sqrt{2\pi m_1 k_B T}} \tag{1.5.9}$$

where m_1 is the mass of a monomer. Therefore the condensation rate is dependent on the supersaturation vapour pressure of the system.

1.5.2.2 Evaporation rate

The evaporation rate, $b(n)$ is not known *a priori*. Instead it is typically found using the detailed balance condition. This takes advantage of the constrained equilibrium state [**19**, **20**] which exists with the same temperature and supersaturation, $\mathcal{S} > 1$ as the original vapour. Crucially, the system is such that the net flux is zero, i.e. $\mathcal{J}_n = 0$ and corresponds to the step before nucleation occurs. It is possible to calculate the evaporation rate by rearranging Eq. (**1.5.7**),

$$b(n+1) = f(n)\frac{\rho_{eq}(n)}{\rho_{eq}(n+1)} \tag{1.5.10}$$

where $\rho_{eq}(n)$ is the n-cluster density within the constrained equilibrium state. It is now possible to rewrite Eq. (**1.5.7**) such that,

$$\mathcal{J}_n \frac{1}{f(n)\rho_{eq}(n)} = \frac{\rho(n)}{\rho_{eq}(n)} - \frac{\rho(n+1)}{\rho_{eq}(n+1)} \tag{1.5.11}$$

1.5.3 Classical nucleation rate

Eq. (**1.5.6**) will rapidly reach a steady state assuming that the monomer population is not exhausted. This steady nonequilibrium state (in the limit $t \to \infty$) will have time independent number densities, $\rho(n)$ which implies that all $\mathcal{J}_n$ are equivalent and can thus be replaced with a steady-state nucleation rate, J. We can now calculate the steady state nucleation rate for a cluster of size N by summing Eq. (**1.5.11**) from $n = 1$ to $n = N$ and cancelling terms on the right hand side gives,

$$J\left[\sum_{n=1}^{N} \frac{1}{f(n)\rho_{eq}(n)}\right] = \frac{\rho(1)}{\rho_{eq}(1)} - \frac{\rho(N+1)}{\rho_{eq}(N+1)} \tag{1.5.12}$$

The population of monomers is assumed to be limitless whereas the nucleation rate is finite which implies the true population of monomers ($\rho(1)$) is comparable to the constrained equilibrium population of monomers ($\rho_{eq}(1)$) and thus[4] $\frac{\rho(1)}{\rho_{eq}(1)} \to 1-$. For large N the constrained state population grows without limit whereas the true population remains finite, thus $\frac{\rho(N)}{\rho_{eq}(N)} \to 0$ as $N \to \infty$. Therefore Eq. (**1.5.12**) can be simplified by choosing a sufficiently large N and rearranging such that,

$$J = \left[\sum_{n=1}^{\infty} \frac{1}{f(n)\rho_{eq}(n)}\right]^{-1} \tag{1.5.13}$$

The standard technique for calculating the right hand side of Eq. (**1.5.13**) is to assume that the sum can be changed into an integral as the major contributions to the sum are for large clusters where the difference in successive terms becomes small. The right hand side can then be rewritten,

$$J = \left[\int_{1}^{\infty} \frac{dn}{f(n)\rho_{eq}(n)}\right]^{-1} \tag{1.5.14}$$

The steepest descent method can be used to evaluate the integral to give an expression for J around the critical cluster n_c such that,

$$J = \mathscr{Z} f(n_c)\rho_{eq}(n_c) \tag{1.5.15}$$

where $\mathscr{Z}$ is known as the Zeldovich factor and is defined as $\mathscr{Z} = \sqrt{-\frac{1}{2\pi}\frac{1}{k_B T}\Delta\varphi''(n_c)}$. It is possible to rewrite Eq. (**1.5.15**) as,

$$J = J_0 \exp\left[-\frac{\Delta\varphi_{i^*}}{k_B T}\right] \tag{1.5.16}$$

and $J_0 = \mathscr{Z} f(n_c)\rho^v$ and ρ^v is the number density of the supersaturated vapour. Eq. (**1.5.16**) can be understood as the rate at which monomers condense upon the critical cluster multiplied by the Zeldovich factor **[21]** which gives the probability that the critical cluster will pass the free energy barrier. The equation along with the capillarity approximation for the calculation of free energy differences gives the full CNT expression for the steady state nucleation rate for the single species system. Its popularity stems from the simple form of the equation and its link to experimentally observable quantities. The dependence of CNT upon accurate free energy calculations is clearly seen from this equation.

[4]This fraction approaches unity from below as the depletion caused by nucleation in the true population implies that $\rho(1) < \rho_{eq}(1)$.

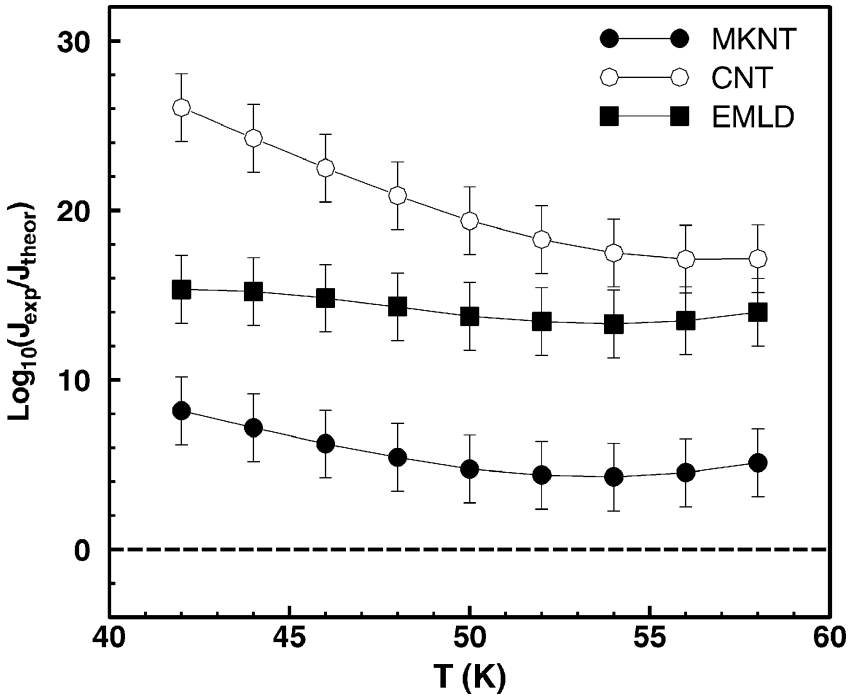


Figure 1.6: Figure 3 from reference [**22**]. Comparison of several theoretical methods against experimental argon nucleation. The theory is shown to give errors of several orders of magnitude. The dashed line represents a perfect agreement. Three theories are compared here which are classical nucleation theory (CNT), mean-field kinetic nucleation theory (MKNT), and extended modified liquid drop model- dynamical nucleation theory (EMLD)

1.5.4 Limitations and extensions

Classical nucleation theory as summarised by Eq. (**1.5.16**) was derived in 1926-1943 [**15**, **21**, **23**]. The success of the theory to describe experimental results is not satisfactory with errors of several orders in magnitude being common [**18**]. Figure **1.6** shows a comparison of models and experiment on argon nucleation made by Kalikmanov *et al.* [**22**], the standard flavour of CNT is found to be of several orders of magnitude in error. Ford [**24**] discusses the shortcomings of CNT at length. The capillarity approximation where the surface tension of a cluster is approximated to be a planar macroscopic quantity is thought to be one of the significant contribution to errors within the CNT calculation.

Several improvements for CNT have been suggested including schemes to include rotational and vibrational degrees of freedom in the free energy [**25**]. Other amendments include enforcing the correct monomer free energy difference [**26**] and is known as internally consistent CNT theory, ICCT [**27**]. The inclusion of curvature contributions to surface energy was investigated by Rao and McMurry [**28**] and found to substantially increase nucleation rates, the effect is most prominent for substances with large surfaces energies and large molecular volumes.

The difficulties involved with CNT have led to an array of alternatives including diffuse interface theory (DIT) [**29**], square gradient theory (SGT) [**30**, **31**], extended liquid drop model with dynamic nucleation theory (ELDM-DNT) [**32**, **33**] and density functional theory[5] [**34**]. All of these theories were analysed in a useful review by Napari et al. [**35**], with the conclusion reached that Density Functional Theory and SGT gave the best performance. The comparisons made are against a theoretical simulation with a Lennard-Jones potential. Each of these methods was found to outperform CNT. A different approach is adopted by Kazil and Lovejoy [**36**] where they treat the problem as a steady state solution to the sources and sinks of aerosols of interest, and parametrize using experimental and theoretical data. Kalikmanov [**37**] developed the mean-field kinetic nucleation theory (MKNT) which made use of a mean field approach and two surface tensions representing planar macroscopic surface tension and microscopic surface tension respectively. This improved the performance of CNT by several orders of magnitude in the case studied shown in Figure **1.6**. This serves to highlight the importance of free energy within the CNT methodology and the responsibility that the capillarity approximation must take in the problems faced by CNT.

Beyond macroscopic theories the use of microscopic techniques can improve upon the assumptions made within CNT. It is possible to use molecular scale simulations to calculate evaporation and condensation rates directly and use these to solve the Becker-Döring equations for the nucleation rate. Alternatively the free energy difference between different clusters can be found via techniques such as the harmonic oscillator approximation (HO) or thermodynamic integration (TI) and can be used to improve the accuracy of the calculated free energies.

[5]The density here refers to the density of molecules and should not be confused with electronic density functional theory studied in Chapter **4** which is based upon the electron density.

Chapter 2

Literature review

In this chapter experimental and theoretical approaches to studying clusters of sulphuric acid and water are reviewed. The theoretical approaches are split between accurate quantum chemistry and efficient empirical methods. In addition reactive empirical approaches are examined and attention is then focused on the empirical valence bond methodology as this is then studied in greater depth in Chapter **6**.

2.1 Experiment

Performing experiments on the phenomena of nucleation is challenging due to the number of parameters which need to be controlled and that the diameter of the clusters of interest are on the scale of nanometres. Doyle [**38**] theorised in 1961 that gaseous mixtures of sulphuric acid and water may undergo a phase change at 1 part per trillion (1ppt), this was subsequently corrected to between 0.1 and 1 part per billion ($0.1 - 1$ ppb) [**39**]. However, it remains that sulphuric acid is one of the most relevant atmospheric nucleation agents [**7**]. One of the earliest experimental studies was performed by Kreidenweis *et al.* [**40**] in 1989, it called into question the accuracy of macroscopic theories (CNT). In 1999 Ball *et al.* [**41**] performed the first experiments where the concentration of sulphuric acid in the new particle being formed was determinable [**7**]. They found at relative humidity[1] (RH) of 15% and at 295 K that the critical cluster of a sulphuric acid and water mixture contained eight sulphuric acid molecules. The addition of a small amount of ammonia, $[NH_3]$ of the order of 10 parts per trillion by volume (pptv) enhances the nucleation rate significantly, and

[1] Relative humidity is defined as the ratio of the partial vapour pressure of gaseous water to the saturated vapour pressure at a defined temperature.

reduces the number of sulphuric acid molecules in the critical cluster. The paper concludes that it is unclear if ammonia is the major species involved in nucleation events close to the earth's surface. The detection limit[2] of the apparatus used was 3 nm which is prone to undercounting[3]. The following year Eisele and Hanson [**42**] found the critical cluster to contain four sulphuric acid molecules at high RH (around 50%) and at 240K which agreed with a bimolecular nucleation theory developed by Kulmala *et al.* [**43**] which parametrised a thermodynamically consistent version of a classical binary homogeneous nucleation theory [**44**]. In 2010 Sipilä *et al.* [**8**] solved an apparent mystery where *in situ* produced gaseous sulphuric acid nucleation rate differed by several orders of magnitude from that of liquid sample production. They found inconsistencies in the sulphuric acid density profiles and inefficiencies with particle counting is sufficient to explain the discrepancy. This has discounted the importance of other sulphur containing species as an important nucleating agent. An ongoing experiment in CERN named CLOUD has recently studied the importance of ammonia and cosmic rays in atmospheric nucleation [**10**]. They observed that ammonia leads to an increase in nucleation rates of sulphuric acid and water of $100 - 1000$. A further increase, between two and tenfold, is seen when the nucleating particles are ionized (provided that the nucleation rate is below that of the ion-pair production rate). A possible source of ionisation events in the atmosphere is interaction with cosmic rays. They concluded that enhancement of nucleation rate is seen with both ammonia and ion addition. However these two factors cannot explain observed boundary layer[4] nucleation[5]. In addition amines have recently been identified as another agent that may play an important role in atmospheric nucleation [**45**]. The CLOUD experiment [**46**] examined the influence of dimethylamine on the sulphuric acid and water system and found that it could raise formation rates $1,000$ fold, theoretically it was suggested that the evaporation rate is repressed by the acid-base interactions within the nucleating cluster.

In conclusion, experiments have shown the importance of sulphuric acid for binary-layer nucleation of water and that sulphuric acid alone does not fully explain observed nucleation rates. Ternary species have been identified such as ammonia which potential play a role. However, as discussed in Section **1.5.4**, CNT has not given satisfactory results when predicting such nucleation rates. The

[2]The detection limit refers to the diameter of the smallest particle that the apparatus can measure.

[3]Undercounting refers to particles which have not grown large enough to be detected by the apparatus and are not counted. Therefore the rate of new particle formation measured is a lower limit on the true value.

[4]The boundary layer is below the troposphere and is characterised by turbulent flow caused by interactions with the surface of the planet.

[5]Boundary-layer nucleation is one source of atmospheric aerosols, see Section **1.2**.

following section will review research looking at theoretical approaches to going beyond the accuracy offered by CNT.

2.2 Theoretical approaches

In parallel to the experimental work described in the previous section there have been efforts to improve upon classical nucleation theory. These effort can be largely grouped into two camps, the first involving complex quantum chemistry approaches which offer high accuracy at a computational expense. The second is the empirical approach which has significant savings in computation which allow for greater statistics to be collected, however the achievable accuracy is compromised. Here literature on both approaches are reviewed.

2.2.1 Quantum chemistry

Geometry optimisation is a process where a configuration under a defined potential is modified such that its potential energy is minimised. The process has become popular in the field of quantum chemistry because calculations are computationally expensive and the accuracy of the geometry is important for the derived properties. Geometry optimisation has proved especially popular when combined with techniques such as: electronic density functional theory, DFT[6] [**47**]; Møller-Plesset theory, MP2[7] [**48**]; and Coupled cluster CC [**49**]. In particular DFT has led to a number of studies of sulphuric acid and water clusters, a review of the DFT methodology is provided in Chapter **4**.

The modern version[8] of DFT was derived in 1964 by Hohenberg and Kohn [**51**] and in 1965 by Kohn and Sham [**52**]. The local density approximation (LDA) was found to work well with bulk metals and became a standard technique in the solid state physics community in the 1970s ([**53**] and references therein). With the second generation of functionals including the generalized gradient approximation (GGA) and hybrid functionals the descriptions of atomistic features greatly improved. This led to DFT being adopted by the chemistry community in the 1990s ([**53**] and references therein). The growing presence of DFT calculations, and especially the considerable effort to develop code packages making it easier to implement DFT and other high level theory made an impact on the

[6]DFT is described in further details in Chapter **4**.
[7]MP2 is a perturbation theory such that the suffixed number refers to the order of the perturbation.
[8]The conceptual origin of electronic DFT is found in the Thomas-Fermi model [**50**] and the Hartree-Fock method [**47**].

aerosol community in the late 1990s. The first calculations performed on hydrated sulphuric acid clusters using DFT were in 1997 by Arstila *et al.* **[54]**. The investigation used the GGA functional BLYP **[55, 56]** to study configurations of a sulphuric acid molecule hydrated with 1 to 3 waters. This study paved the way for the use of quantum chemistry techniques to be used in the study of the nucleation and chemistry of the sulphuric acid and water system. The main conclusion revealed inconsistencies between that of the quantum chemistry approach and a classical theory for the Gibbs free energy given by Jaecker-Voirol **[57]**. They found that a level of hydration of three to four waters was sufficient to allow proton transfer to occur between the sulphuric acid and the water network. The nature of deprotonation event is of importance as an ionised cluster is thought to have an enhanced nucleation rate in comparison to a neutral cluster **[46]**.

This was followed by extensive work from two groups, Ianni and Bandy (1998 **[58]** and 1999 **[59]**) and Re *et al.* (1999 **[60]**). The first paper by Ianni and Bandy focussed on the question of the equilibrium state of $[H_2SO_4] \cdot [H_2O]_n$ for $n = 1 - 7$ using the hybrid B3LYP functional[9]. They found that the structure of small clusters of water and sulphuric acid differs from that found in bulk solution. The study concludes that the globally stable configuration is neutral for the first six waters, becoming ionized after seven waters. This contradicted the results of Arstila *et al.* **[54]** from the previous year. The following year Ianni and Bandy published a second paper **[59]** which extended the work by looking at $[H_2SO_4]_2 \cdot n[H_2O]$ clusters for $n = 0 - 6$ again using the B3LYP functional. The paper concluded that there was a kinetic barrier in forming $[H_2SO_4]_2 \cdot n[H_2O]$ clusters, and suggested several methods by which the barrier can be overcome. It is interesting to note that the study did not consider the ionized system to be important. The same year Re *et al.* **[60]** investigated $[H_2SO_4] \cdot [H_2O]_n$ for $n = 1 - 5$ and significantly extended the number of configurations studied by Ianni and Bandy the previous year **[58]**. The configurations identified for the trihydrated sulphuric acid are shown in Figure **2.1**. They concluded that several configurations are expected to coexist in these clusters which include both neutral and ionized states. They also performed Infra-red (IR) spectra calculations for experimental comparison. These studies form a useful data set of geometry optimised configurations and relative energies of small clusters of sulphuric acid and water.

In addition there have been many other DFT level geometry optimisation studies which are

[9]The B3LYP functional is based upon the BLYP functional where the exchange energy is combination of exact Hartree-Fock exchange and exchange introduced by the BLYP functional. The 3 refers to the introduction of three parameters compared to that of BLYP, these parameters control the mixing of the two exchange energies. It was developed by Becke **[61]**.

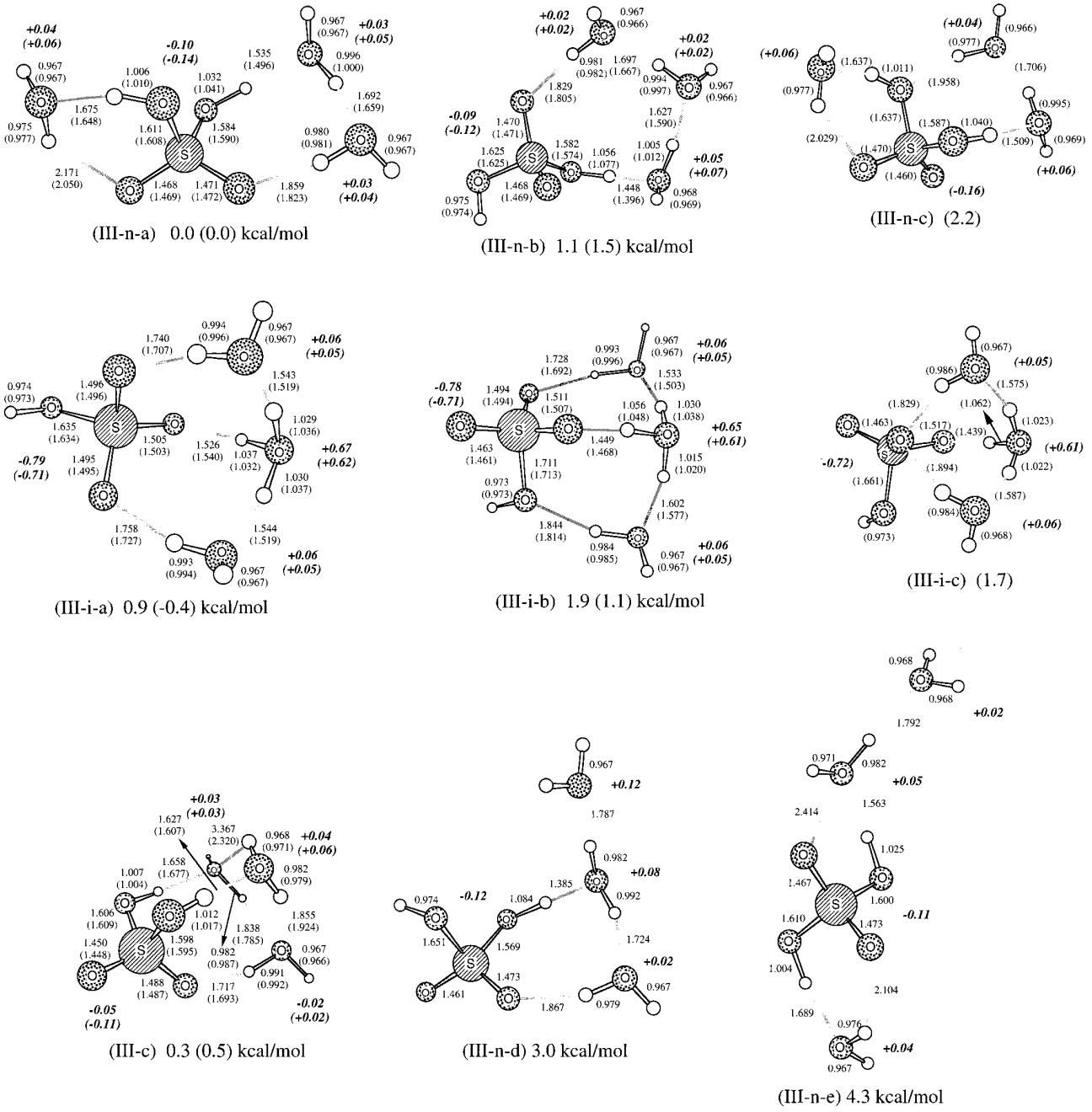


Figure 2.1: Image displayed is that of Figure 3 in Reference [**60**]. Several geometry optimised configurations are detailed for a cluster composed of one sulphuric acid and three water molecules. References [**54**, **58**–**60**] contain many similar details for small clusters of sulphuric acid and water.

noteworthy when studying small clusters of sulphuric acid and water **[62–72]**. One conclusion of these papers was that $[SO_3] \cdot [H_2O]_4 \rightarrow [H_2SO_4] \cdot [H_2O]_3$ was barrier-less **[66]** using the B3LYP functional and using the higher level MP2 theory. In 2003 Ding *et al.* **[68]** considered systems involving two sulphuric acid molecules and zero to eight water molecules using the BLYP functional. They found that the first deprotonation event occurred after two waters are added to the two sulphuric acid system. At four waters it is more favourable for both sulphuric acids to have undergone the first deprotonation event. The second deprotonation event to form the sulphate ion ($[SO_4]^{-2}$) is not seen. They suggested that two waters per sulphuric acid are needed for each sulphuric acid to undergo the first ionisation event. An extensive study on $[H_2SO_4] \cdot [H_2O]_n$ for $n = 1 - 3$ was performed by Natsheh *et al.* **[67]** in 2004. This study compared six functionals and concluded that the PW91 functional[10] models the sulphuric acid molecule more faithfully than the B3LYP for nonbonded interactions **[74]**. The PW91 functional was then used to calculate the vibrational frequencies, dipole moments and the bonding energies of the clusters. Also in 2004 Arrouvel *et al.* **[65]** studied clusters and crystals of sulphuric acid and water using the B3LYP and PW91 functionals. Their main conclusion was that for single sulphuric acid molecules three or more waters were required to ionise the sulphuric acid. When there are two sulphuric acid molecules present in the cluster then two waters per sulphuric acid leads to a stable configuration where both sulphuric acid molecules have ionised once. These results are in agreement with earlier results obtained by Ding *et al.* **[68]**. It was also found that in the crystal structure the acid was always fully ionised to form a sulphate ion. Ding and Laasonen **[64]** investigated clusters of $[H_2SO_4] \cdot [H_2O]_n$ for $n = 6 - 9$ in 2004 using the BLYP and PW91 functionals and MP2 (with the resolution of identity technique (RI) **[75, 76]**). They concluded that eight waters for one sulphuric acid molecule is the minimum required to fully ionise one sulphuric acid and that in most small clusters the species is not an important consideration. Nadykto *et al.* **[63]** investigated the sign preference in sulphuric acid and water nucleation using the PW91 functional and concluded that the hydration of sulphuric acid favoured cations while the affinity to other sulphuric acid molecules is higher for positive ions, which are competing aims. This study was followed up a year later by Kurtén *et al.* **[62]** which concluded that sulphuric acid binds more strongly to anions than to cations. It also suggests that the preference of molecules can be predicted using general chemical considerations to a qualitative level. This study used the BLYP

[10]The PW91 functional is of the GGA variety and was developed by Perdew and Wang **[73]**.

functional and the MP2-RI technique. In 2012 Herb *et al.* **[72]** investigated large clusters of sulphuric acid, bisulphate and water with and without ammonia, $[NH_3]$ using the PW91 functional. They concluded that the presence of ammonia is favourable for cluster growth and that the affinity of ammonia to the cluster is related to the concentration of sulphuric acid in the cluster and suggest that this trend will continue for larger clusters.

In 2012 Temelso *et al.* **[69]** studied small clusters of sulphuric acid and water using RI-MP2 methodology. They identified several configurations of the cluster $[H_2SO_4(H_2O)_n]$ for $n = 1 - 6$. They suggest that ionisation becomes favoured at a hydration of 3 or more water molecules and that the formation of clusters composed of sulphuric acid and five or more waters is unfavourable at temperatures greater than 273.15 K. They also suggest that the critical cluster for sulphuric acid and water must contain more than one sulphuric acid, in agreement with experiment **[7]**. They conclude that it is unlikely that binary sulphuric acid and water nucleation is a significant contributor to new particle formation at ambient conditions in the continental boundary layer, however they suggest that binary sulphuric acid and water nucleation may contribute in colder troposphere regions. Figure **2.2** shows the expected sulphuric acid and water cluster equilibrium distributions in atmospherically relevant conditions, the figure shows that clusters of $1 - 3$ waters are present, however the larger clusters do not show a sizeable population. The same group later also studied hydrated bisulphate **[70]** and the hydrated sulphuric acid dimer clusters **[71]**. They concluded that the bisulphate ion by itself cannot lead to the nucleation of water, however, its strong binding energy with sulphuric acid makes it a possible agent for ion induced nucleation. They suggest that four water molecules are required for the each sulphuric acid in the dimer to undergo its first ionisation event and that the addition of monohydrated sulphuric acid to form the dimer cluster is more favourable than that of the sulphuric acid molecule.

The studies presented in this section have found several key features of small clusters of sulphuric acid and water. The first deprotonation occurs in the trihydrated sulphuric acid **[65, 69]**. A lower level of hydration is required in clusters involving two sulphuric acids **[68]**. A large number of geometry optimised clusters have been identified and the free energy of the clusters have been calculated using the harmonic approximation. However the relative importance of different configurations (of the same cluster) is still an open question. Dynamics offers one approach to answering this question as it samples configurational space.

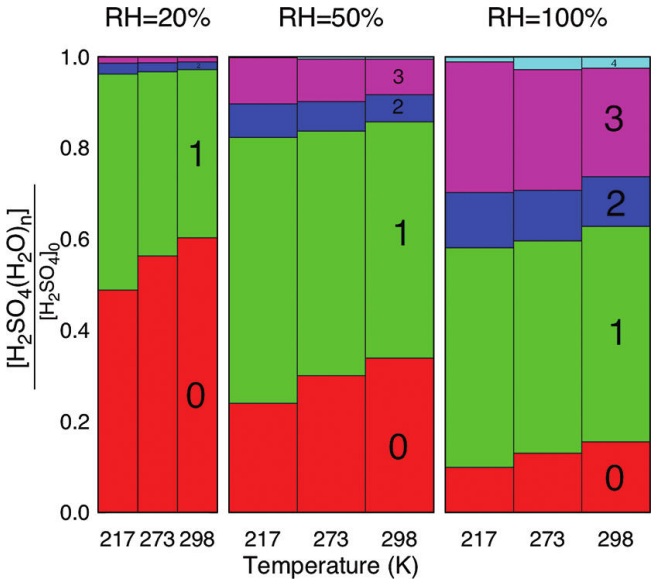


Figure 2.2: Image showing the expected equilibrium populations of clusters which are hydrated according to the number shown in the columns divided by the monomer population of sulphuric acid molecules. The data was obtained assuming a saturation (100% Relative humidity, RH) vapour pressure of $[H_2O] = 9.9 \times 10^{14}$, 1.6×10^{17} and $7.7 \times 10^{17} cm^{-3}$ at temperature of 216.65, 273.15 and 298.15 K respectively with a monomer sulphuric acid density of $5 \times 10^7 cm^{-3}$. As the RH is increased the monomer population is diminished as more clusters of higher hydration are formed. The figure displayed here is that of Figure 12 in reference **[69]**.

2.2.2 Molecular dynamics with quantum chemistry

The use of DFT within a molecular dynamics scheme was pioneered in the 1980s with techniques such as the Car and Parrinello method **[77]** and the Born-Oppenheimer MD (BOMD). The BOMD method evaluates the electronic density at each time-step based on the current nuclei configuration (see Figure **4.1**). Investigations of sulphuric acid and water systems using DFT MD have been limited by the computational expense of performing the simulations, however efforts have been made recently due to increasing computational power available. The first study was in 2007 by Choe *et al.* **[78]** which used the PBE functional[11] with the BOMD method to look at aqueous sulphuric acid at 298 K. They concluded that the diffusion coefficient for the proton is reduced in the 1:3[12] system rather than the 1:64. This hinders the proton shuffling mechanism (known as the Grotthuss mechanism **[80]**). The following year Anderson *et al.* **[81]** studied the formation of $[H_2SO_4]_m \cdot \text{Base} \cdot [H_2O]_6$ for $m = 1 - 2$ using the BLYP functional at a temperature of 250K with BOMD. They conclude that the simulation indicates the importance of two sulphuric acid molecules for the formation of double ions ($[SO_4]^{2-}$)

[11] The functional is of the GGA variety and takes it name from its three authors, Perdew-Burke-Ernzerhof **[79]**.
[12] Where n:m represents m water molecules for every n sulphuric acid molecules in the system.

and that further simulation is needed to investigate the free energy landscape of such systems. In the same year Hammerich *et al.* [**82**] investigated a single sulphuric acid in a periodic box with 63 water molecules. This was simulated at $320 - 326$K with the BLYP and the HCTH/120 functional[13]. The BOMD scheme for performing MD is used. They concluded that the ionization observed had a qualitative agreement with experimental (limited by the simulation time of the order of 10 ps). They also argued that an order of magnitude greater in simulation time was needed to observe events leading to the second ionisation event occurring.

2.2.3 Path integral molecular dynamics

The quantum chemistry techniques describe in the previous two sections treat the electrons in the system as quantum particles via a wavefunction representation. The nuclei, however, are treated as classical point like particles. In some systems this assumption is questionable, it is useful to be able to probe the importance of quantum nuclear dynamics. Path integral molecule dynamics is such a method. It is based on Feynman Path Integrals view of quantum dynamics [**88**]. An excellent review of PIMD is found in chapter 12 of "Statistical Mechanics: Theory and Molecular simulation" by Tuckerman [**12**] and is reviewed in Chapter **5**. The method is used to add quantum dynamics to the nuclei in the DFT simulation by representing the particle as a series of connected beads. Usually only the lightest elements in the periodic table are considered for PIMD simulation due to the expense of performing simulations and that the effects are expected to be most prevalent in the lightest nuclei at temperatures where the potential energy of the system dominates over the kinetic energy. In the sulphuric acid and water system only the hydrogen atoms and in particular those involved in hydrogen bonds are thought to have significant contribution from such quantum dynamics.

The author is only aware of one group currently working on sulphuric acid and water clusters using a PIMD technique. Kakizaki *et al.* [**89**] parametrised the PM6 model [**90**] which in itself is a parametrised version of the Neglect of Diatomic Differential Overlap (NDDO) method [**91**]. They parametrized to data from references [**54**, **58**, **60**, **78**]. The model found that acid dissociation increased with cluster size as expected. The same group followed up with a study on the first and second deprotonation events [**92**] using the PIMD PM6 described in the first publication. They concluded

[13]The HCTH functional is of the GGA variety, its name is derived from its four authors, Hamprecht-Cohen-Tozer-Handy and the number of molecules to which the functional was fitted, in this case it was 120 [**83**].

Method	Advantages	Disadvantages
GAP [**84**]	The method is very adaptable and does not depend upon an underlying empirical model.	It is a new method and as such it not well tested. It is not a trivial method to understand and implement.
ReaxFF [**85**]	The method scales at the same rate as empirical models and is an established simulation tool in MD codes.	It uses the bond order methodology. After discussions with past users of the method it became clear that parametrisation would not be straightforward.
QM/MM [**86**]	QM/MM allows the full power of quantum chemistry methods while allowing for large simulations based on the efficiency of empirical potentials. It has been implemented in available MD codes.	QM/MM method requires specific regions to be treated as either quantum or classical. This affects the flexibility of the model.
EVB [**87**]	The method relies on empirical potentials which are well understood and requires relatively little additional work.	Its reliance on empirical potentials could lead to issues with transferability. The model is dependent on the form of a second empirical term which increases the number of parameters to fit.

Table 2.1: Table discusses the advantages and disadvantages of several reactive schemes.

that four waters are needed for the reaction $[H_2SO_4] \rightarrow [HSO_4]^-$ to occur and ten to twelve for $[HSO_4]^- \rightarrow [SO_4]^{2-}$ to occur. In additional the coordination number for the proton-accepting oxygen is found to be important in the proton exchange process.

2.3 Empirical potentials for sulphuric acid and water

Several classical schemes have been used to study sulphuric acid and water clusters. In 1998 Kusaka *et al.* [**94**] developed a grand canonical Monte Carlo model based on rigid molecules and concluded that clusters are highly non-spherical and that bulk-like behaviour only emerges when there are at least 240 water molecules and $1-3$ molecules of either sulphuric acid or bisulphate in the cluster. Kathmann and Hale [**95**] presented an innovative approach which included a rigid water and sulphate structure and a free $H^{\delta+}$ ion simulated with a Monte Carlo approach. Ding *et al.* [**96**] developed a flexible model for sulphuric acid, bisulphate, hydronium and water species. The ability to transfer

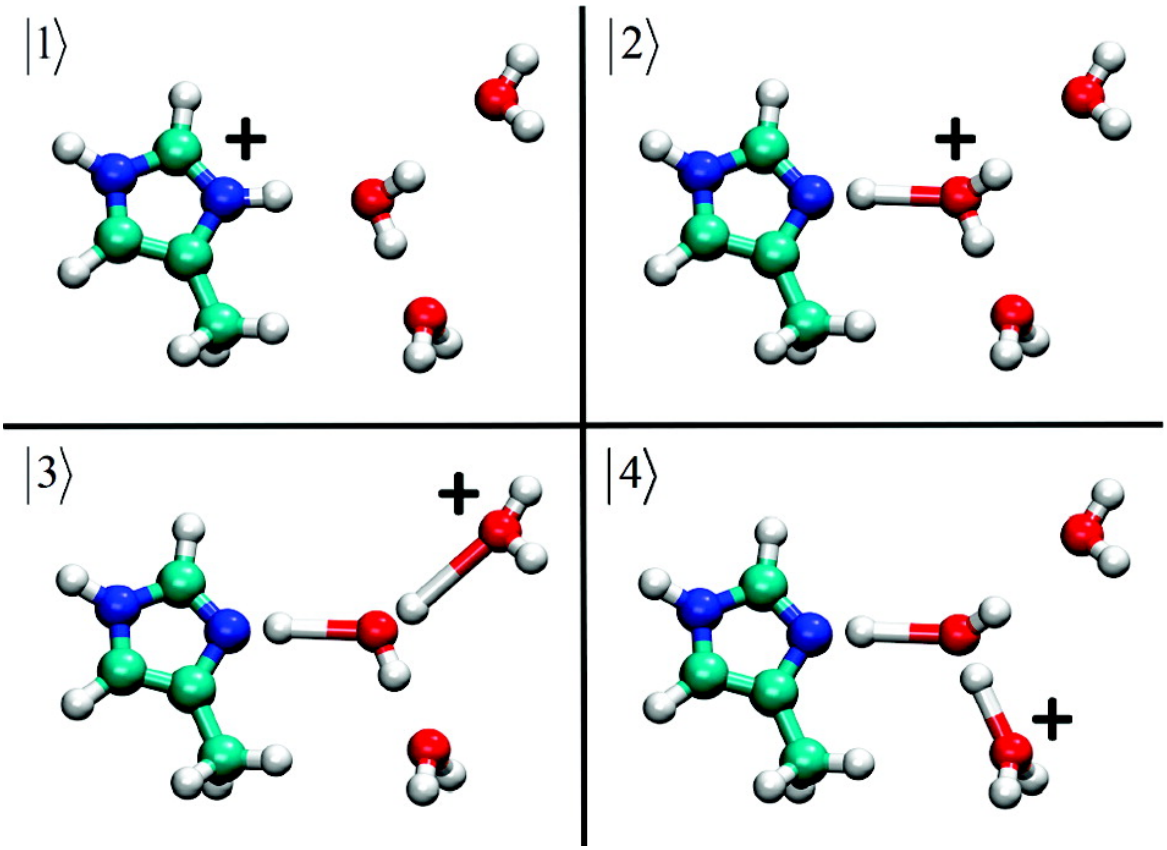


Figure 2.3: EVB diabatic states for 4-Methylimidazole and water, image is Figure 1 in reference **[93]**.

protons was absent from the model. Toivola *et al.* **[97]** used the model to study structural details of a planar liquid-vapour interface using clusters of 2000 molecules. They found that when the sulphuric acid mole fraction is less than 0.1 the acid lies at the cluster surface and that the structure of the clusters is strongly dependent on the number of bisulphates present.

2.4 Reactive potentials

A reactive empirical procedure is based on the ability to break and form bonds, the difficulties in accurately representing these reactions has prevented this feature being present in the majority of molecular dynamic schemes. However the ability to proton transfer is a key characteristic of the sulphuric acid and water system and its neglect would be questionable. It was decided in this project to develop such a reactive potential specifically for the sulphuric acid and water system. Here readily available reactive schemes are reviewed and an appropriate method is chosen for the project.

2.4.1 Reactive empirical potentials

There are several schemes available which allow reactions to occur within an empirical framework including: the EVB model [**87**] based on a superposition of underlying empirical bonding states of the system (see Figure **2.3** for an example of these EVB states), the Gaussian approximation potential (GAP) [**84**] which constructs a potential energy surface by using a data set and the Gaussian process regression method to interpolate between these data points, and ReaxFF [**85**] which uses the bond order methodology. In addition there are so called quantum mechanics/molecular mechanics, QM/MM techniques (as introduced by Warshel and Levitt [**98**] and reviewed by Lin and Truhlar [**86**]) which perform classical simulations in which particular regions are treated quantum mechanically. Table **2.1** discusses some advantages and disadvantages for these methods. The assessment of the models fell on the speed at which it could be implemented due to time constraints of the Ph.D. project. For this reason the EVB model was deemed the most straightforward to implement due to it's similarity to more tradition non-reactive empirical methods.

2.4.2 The empirical valence bond method

The EVB methodology has its origins in the 1980s when Warshel and Weiss [**99**] extended the valence bond model to include an empirical evaluation of environmental effects. The EVB method was reviewed by Kamerlin and Warshel [**87**]. Progress was made in 1998 by Vuilleumier and Borgis [**100**] who presented the extended EVB model and Schmitt and Voth [**101**] who presented the multistate EVB, MS-EVB model. Both of these models had the capacity to simulate one excess proton in a water network. The MS-EVB model was then refined into the MS-EVB2 model in 2002 by Day *et al.* [**102**] and into the MS-EVB3 model in 2008 by Wu *et al.* [**103**]. The MS-EVB3 model was used as a framework from which to build an MS-EVB model for sulphuric acid and water system.

The MS-EVB model has been implemented in a number of ways. In 2001 Čuma *et al.* [**104**] used it to model a system of an imidazolium cation, $[ImH]^+$ in an aqueous solution to study the ionisation of a weak acid using the MS-EVB approach. They found that the multiple states simulated by the model are important for the accuracy of the potential energy barrier of ionisation for the acid. The water network surrounding the acid is found to be localised when the acid is charged in contrast to being delocalised when the acid is neutral. In 2006 Maupin *et al.* [**105**] studied amino acids using a modified

MS-EVB model. The model focused on the addition of histidine and glutamic acids to a water system. Each system contained one excess proton which is associated to the neutral acid structure. They develop a general parametrisation scheme which is used for finding the EVB parameters for histidine and glutamic acid. The paper concludes with the observation that several expected features are present such as Grotthuss shuffling and formation of Zundel ($[H_5O_2]^+$) and Eigen ($[H_9O_4]^+$) cations. It is suggested that the model is to be used to study proton transport pathways in biological systems. In 2011 the same group used the framework to study a proton transport event involved in the catalyst carbonic anhydrase II **[93]**. The 4-Methylimidazole molecule was parametrised for this work with the EVB states shown in Figure **2.3**. They conclude by suggesting possible pathways for the catalyst, and that the 4MI molecule can act as either a proton donor or as an intermediate species in this system. Extensions to the MS-EVB model have included using an anharmonic underlying classical potential as in the aMS-EVB3 model developed by Park *et al.* **[106]** and the use of polarized classical model as presented by Brancato and Tuckerman **[107]**.

The MS-EVB model has been shown to be a reliable method for introducing reactions to a system in an efficient way. However questions remain about the empirical representation of the reactions, and efficiency when simulating several reactions simultaneously. The model is independent of the system and so it is feasible that a code could be developed that would dramatically reduce the effort need to implement MS-EVB for an arbitrary system, this is explored further in Section **9.3**.

Chapter 3

Molecular Dynamics

This chapter gives an overview of molecular dynamics and several useful techniques that complement molecular dynamic simulations. There are several useful texts on implementing and understanding molecular dynamic simulations [**108–110**].

3.1 Ensembles and the ergodic hypothesis

An ensemble describes a system at equilibrium which has well defined statistical properties. The microcanonical ensemble is a commonly used ensemble which describes an isolated system and has a constant total system energy, E, volume and temperature. The microcanonical ensemble has entropy [**109**] as defined by the Boltzmann entropy,

$$S = k_B \ln \Omega(NVE) \tag{3.1.1}$$

where $\Omega(NVE)$ is the number of microstates corresponding with the observed macrostate. The canonical system represents an isolated system that is connected to a heat bath. The free energy which is minimised by the ensemble is that of the Helmholtz free energy, $\mathcal{F}$ (see Table **1.1**). Several commonly used ensembles for simulations are listed in Table **3.1**. Understanding the free energy of the system is vital for calculating the properties of nucleation (see Section **1.5**).

Ensemble name	**Conserved quantities**	**Free energy**
Canonical	NVT	$\mathcal{F}$ (Helmholtz)
Isothermal-isobaric	NPT	$\mathcal{G}$ (Gibbs)

Table 3.1: Properties of some commonly used ensembles.

An important concept in performing microscopic NVE simulations is the ergodic hypothesis [**109**] which states that over a suitably long time the system will traverse each available microstate with an equal probability. One consequence of the ergodic hypothesis is that the time average of a property is equivalent to the ensemble average. Caution needs to be used in respect to the appropriate duration of simulations. Typically the relaxation period[1] of the property of interest gives a suitable time scale for the simulations. The next section describes two schemes for performing the calculations necessary for determining the property of interest.

3.2 Molecular dynamics and Monte Carlo

When studying systems at the microscopic scale we are interested in certain properties which require the evaluation of multidimensional integrals over the possible microstates of the system. There are two widely used schemes for evaluating such integrals, namely Molecular Dynamics, MD, and Monte Carlo, MC [**109**]. MC samples phase space in a purely stochastic manner. This is achieved by moving the system by a random amount in phase space and then applying a set of rules to determine if the movement is accepted or rejected, this process is repeated many times in a typical MC simulation. For a canonical ensemble the average of property $A(\mathbf{r}^N, \mathbf{p}^N)$ is found by evaluating the following integral,

$$\left\langle A(\mathbf{r}^N, \mathbf{p}^N) \right\rangle = \frac{1}{\mathcal{Z}} \int \dots \int \exp\left[-\beta \mathscr{H}(\mathbf{r}^N, \mathbf{p}^N)\right] A(\mathbf{r}^N, \mathbf{p}^N) d\mathbf{r}^N d\mathbf{p}^N \tag{3.2.1}$$

where $\mathbf{r}^N$ and $\mathbf{p}^N$ are the position and momentum vectors of a system of N particles. β is the inverse temperature and is defined by $\beta = (k_B T)^{-1}$. $\mathscr{H}$ is the classical Hamiltonian[2] defined by $\mathscr{H}(\mathbf{r}^N, \mathbf{p}^N) = \sum_i p_i^2/2m_i + \mathscr{U}(\mathbf{r}^N)$ with the potential energy, $\mathscr{U}(\mathbf{r}^N)$. m_i and p_i are the mass and the momentum of particle i. $\mathcal{Z}$ is the canonical partition function and is defined by,

$$\mathcal{Z} = \int \dots \int \exp\left[-\beta \mathscr{H}(\mathbf{r}^N, \mathbf{p}^N)\right] d\mathbf{r}^N d\mathbf{p}^N \tag{3.2.2}$$

The ensemble average $\left\langle A(\mathbf{r}^N, \mathbf{p}^N) \right\rangle$ is estimated by taking the average of the property over the

[1] The correct relaxation period for the system can be estimated when $\sigma / \langle A \rangle$ is of order N where $\langle A \rangle$ is the average of the extensive (that is of order N) property, A, and σ is the standard deviation of A [**109**]. One example is that of temperature which is detailed in Section **A.1**.

[2] For convenience the system is defined as being classical, however MD and MC can equally be applied to non-classical simulations such as density functional theory as described in Chapters **4** and **5**.

course of the MC simulation and equating this to quantity to $\langle A(\mathbf{r}^N, \mathbf{p}^N)\rangle$ via the assumption of equal a priori probability[3]. In order to rely on the estimated averages we must be sure that a sufficient amount of sampling has taken place during the MC simulation. This is a significant task and there are several techniques used to reduce the amount of computational work required. The modern version of the Monte Carlo technique was introduced by Metropolis and Ulam [**111**] in 1949 where the potential energy of the system, $\mathscr{U}(\mathbf{r}^N)$ was employed for the purpose of sample biasing to drastically reduce the computation work required. This is justified under the assumption that the Boltzmann factor, $\exp\left[-\beta\mathscr{U}(\mathbf{r}^N)\right]$, results in a large section of phase space having a negligible contribution to Eq. (**3.2.1**). More information on the Monte Carlo method can be found in the book by Kalos and Whitlock [**112**].

In contrast to the stochastic method of Monte Carlo, Molecular Dynamics is deterministic[4] as each sampled configuration in the simulation is related to the initial configuration via an equation of motion such as Newton's:

$$\mathbf{F}_i(t) = m_i\ddot{\mathbf{r}}_i(t) = -\frac{\partial\mathscr{U}(\mathbf{r}^N)}{\partial\mathbf{r}_i} \tag{3.2.3}$$

where $\mathbf{F}_i$ is the force vector acting upon particle i and m_i is the mass of particle i. A numerical approach is typically used to evaluate Eq. (**3.2.3**) due to difficulties in obtaining an analytical solution when the form of $\mathscr{U}$ is non-trivial. An example of such a numerical integration scheme for Eq. (**3.2.3**) is the Velocity Verlet algorithm which is discussed in the next section.

For molecular Dynamics the ensemble averaged property, $\langle A\rangle$, are estimated by calculating the time averaged value of the property of interest, i.e.

$$\langle A\rangle = \lim_{t\to\infty}\frac{1}{t}\int_{t_0}^{t_0+t} A(\tau)d\tau \tag{3.2.4}$$

where t is the length of the simulation and τ is a dummy variable used to integrate over the simulation. The ergodic hypothesis is then used to equate this to the ensemble average, this reasoning relies on the assumption that a sufficient amount of statistics has been gathered.

[3]Equal a priori postulates that the system has equal probablity of visiting every microstate available to it.

[4]The use of certain thermostats in a MD simulation can result in it not being deterministic, the Langevin thermostat is one such thermostat and is described in Section **3.5**.

3.3 Velocity Verlet algorithm

A key component of MD is a viable scheme for numerically integrating the system over a select time period. The Velocity Verlet algorithm [**113**, **114**] is a popular example of such a procedure [**109**] and is known as a third order Störmer algorithm. The derivation starts by performing a Taylor series expansion of position, x at time, $t + \Delta t$ to the second order:

$$x(t + \Delta t) = x(t) + v(t)\Delta t + \frac{1}{2m_i}F(t)\Delta t^2 + \mathcal{O}(\Delta t^3) \tag{3.3.1}$$

where v and F are the velocity and force respectively. m_i is the mass of the particle. The next stage in the derivation involves performing the reverse operation, that is to expand around position $x(t + \Delta t)$ at time t to find,

$$x(t) = x(t + \Delta t) - v(t + \Delta t)\Delta t + \frac{1}{2m_i}F(t + \Delta t)\Delta t^2 + \mathcal{O}(\Delta t^3) \tag{3.3.2}$$

By substituting Eq. (**3.3.2**) into Eq. (**3.3.1**) and solving for $v(t + \Delta t)$ gives,

$$v(t + \Delta t) = v(t) + \frac{\Delta t^2}{2m_i}\left(F(t) + F(t + \Delta t)\right) + \mathcal{O}(\Delta t^3) \tag{3.3.3}$$

The Velocity Verlet algorithm for evolving a system can therefore be expressed as iteratively performing equations (**3.3.1**) and (**3.3.3**).

3.4 Thermostats

Several schemes have been developed to control the expected temperature statistics of Molecular Dynamic schemes, these are known as therostats. Popular thermostats include the Nosé-Hoover [**115**, **116**] which introduces an extra degree of freedom which is associated with the heat bath, and the velocity scaling Berendsen [**117**]. However for this project the stochastic Langevin thermostat [**12**] was found to be the most suitable as it produced Maxwell-Boltzmann distributions for the velocity of each individual atom and the expected standard deviation of the temperature of system as expected (see Appendix **A**). The Langevin thermostat introduces a noise term and a friction term to the equation of motion. A useful review of Langevin dynamics is provided in section 15 of the *Statistical Mechanics:*

Theory and Molecular Simulation book by Tuckerman [**12**]. The following section describes an issue found when applying a constraint within the simulation that uses the Langevin thermostat.

3.5 The Langevin thermostat and small systems

3.5.1 Introduction

This section resolves some peculiar behaviour encountered when using the DL_POLY (version 4.03) code [**118**] to perform Langevin thermostated [**12**] MD simulations of small clusters composed of sulphuric acid molecules. The temperature control imposed by the implementation of a Langevin thermostat was found to deviate from expectation: groups of atoms were brought to different temperatures depending on their mass. Such species-dependent temperatures were deduced by fitting Maxwell-Boltzmann distributions (see Appendix **A**) to velocity distributions of atoms, grouped by mass, extracted from 'equilibrated' system trajectories. Further investigation revealed that this discrepancy was a function of the number of particles, N, in the system as shown in Figure **3.1**. Clearly this is not the desired outcome.

It is shown that this behaviour is a direct result of a constraint imposed in the code whereby no thermal noise acts upon the centre of mass (CoM). It is desirable to fix the centre of mass of a system in simulation as translation of the entire system is often of little importance to that being studied. In Section **3.5.2** the expected temperatures of atomic species generated by such a constrained Langevin thermostat is derived. Various numerical tests are performed to validate the analysis and the results are discussed in Section **3.5.3**.

3.5.2 Theory

3.5.2.1 Langevin derivation of thermalisation temperature

Langevin thermostating of a set of non-interacting particles moving in one spatial dimension is implemented through the use of the equation of motion [**12**],

$$m_i\dot{v}_i = -\gamma m_i v_i + b_i m_i \xi_i(t) \tag{3.5.1}$$

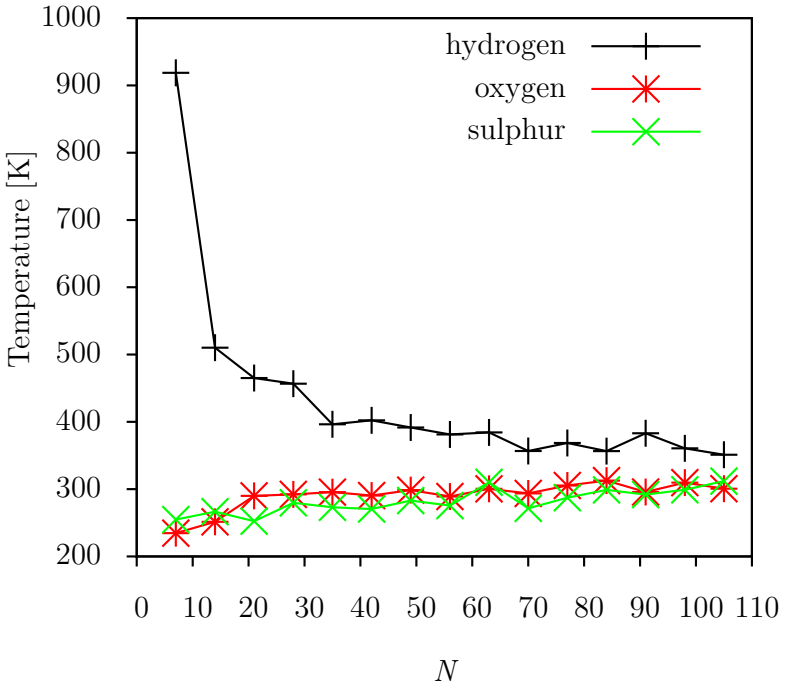


Figure 3.1: Dependence of the temperature of each atomic species against the number of particles, N, in a simulation of up to 15 sulphuric acid molecules in a Langevin thermostated DL_POLY_4.03 simulation with a target temperature of 300 K. The heavier atoms (oxygen and sulphur) are seen to be cooler than desired and the lighter atoms (hydrogen) are hotter.

where m_i and v_i are the mass and velocity for particle i, γ is the friction coefficient and $b_i = \sqrt{2\gamma k_B T/m_i}$ where k_B is the Boltzmann constant and T is the target temperature. $\xi_i(t)$ is a noise term with the following statistical properties,

$$\langle \xi_i(t) \rangle = 0$$

$$\langle \xi_i(t)\xi_i(t') \rangle = \delta(t - t')$$

$$\langle \xi_i(t)\xi_j(t') \rangle_{i \neq j} = 0$$

Mutual interactions could be added to these equations, but this should not affect the thermalisation, and are not considered for simplicity. However, in DL_POLY_4.03 the sum of all the Langevin noise terms is constrained to be zero. This is perhaps motivated by a desire not to allow the thermostat to disturb the CoM motion of the system; in the scheme the CoM, momentum is allowed to relax deterministically towards zero. However, this constraint affects the function of the thermostat and leads to undesirable dynamical behaviour as shown below.

The derivation of the effect such a constraint has on the thermostat is started by defining C as

the sum of the noise terms $\sum_1^N b_j m_j \xi_j(t)$. The random force on the CoM may then be eliminated by subtracting C/N from the equation of motion (Eq. (**3.5.1**)) of each particle. After dividing by m_i we obtain,

$$\dot{v}_i = -\gamma v_i + \frac{N-1}{N} b_i \xi_i(t) - \frac{1}{N} \sum_{j=1, j\neq i}^{N} b_j \frac{m_j}{m_i} \xi_j(t) \tag{3.5.2}$$

and a standard solution to such a Langevin equation with several independent noise terms [**12**] gives,

$$\begin{aligned} v_i(t) = v_i(0)\mathrm{e}^{-\gamma t} &+ \frac{N-1}{N} b_i \int_0^t \mathrm{e}^{-\gamma(t-t')} \xi_i(t') dt' \\ &- \frac{1}{N} \sum_{j=1, j\neq i}^{N} b_j \frac{m_j}{m_i} \int_0^t \mathrm{e}^{-\gamma(t-t')} \xi_j(t') dt' \end{aligned} \tag{3.5.3}$$

from which we can obtain $v_i(t)^2$,

$$\begin{aligned} v_i^2(t) = v_i(0)^2 \mathrm{e}^{-2\gamma t} &+ \left(\frac{N-1}{N}\right)^2 b_i^2 \int_0^t \int_0^t \mathrm{e}^{-\gamma(t-t'')} \mathrm{e}^{-\gamma(t-t')} \xi_i(t') \xi_i(t'') dt' dt'' \\ &+ \left(\frac{1}{N}\right)^2 \sum_{j=1, j\neq i}^{N} \sum_{k=1, k\neq i}^{N} b_j b_k \frac{m_j m_k}{m_i^2} \int_0^t \int_0^t \mathrm{e}^{-\gamma(t-t'')} \mathrm{e}^{-\gamma(t-t')} \xi_i(t') \xi_i(t'') dt' dt'' \\ &+ v_i(0)\mathrm{e}^{-\gamma t} \frac{N-1}{N} b_i \int_0^t \mathrm{e}^{-\gamma(t-t')} \xi_i(t') dt' \\ &- v_i(0)\mathrm{e}^{-\gamma t} \frac{1}{N} \sum_{j=1, j\neq i}^{N} b_j \frac{m_j}{m_i} \int_0^t \mathrm{e}^{-\gamma(t-t')} \xi_j(t') dt' \\ &- \frac{N-1}{N^2} b_i \sum_{j=1, j\neq i}^{N} b_j \frac{m_j}{m_i} \int_0^t \int_0^t \mathrm{e}^{-\gamma(t-t'')} \mathrm{e}^{-\gamma(t-t')} \xi_i(t') \xi_j(t'') dt' dt'' \end{aligned} \tag{3.5.4}$$

taking the average of $v_i^2(t)$ and using the properties of $\xi_i(t)$ and that $\langle \xi_j(t > 0) v_i(0) \rangle = 0$ (i.e. the force introduced by $\xi_i(t)$ is not correlated with the initial velocity) to eliminate the cross terms we get,

$$\begin{aligned} \langle v_i^2(t) \rangle = v_i(0)^2 \mathrm{e}^{-2\gamma t} &+ \left(\frac{N-1}{N}\right)^2 b_i^2 \int_0^t \int_0^t \mathrm{e}^{-\gamma(t-t'')} \mathrm{e}^{-\gamma(t-t')} \langle \xi_i(t') \xi_i(t'') \rangle dt' dt'' \\ &+ \left(\frac{1}{N}\right)^2 \sum_{j=1, j\neq i}^{N} \left(b_j \frac{m_j}{m_i}\right)^2 \int_0^t \int_0^t \mathrm{e}^{-\gamma(t-t'')} \mathrm{e}^{-\gamma(t-t')} \langle \xi_i(t') \xi_i(t'') \rangle dt' dt'' \end{aligned} \tag{3.5.5}$$

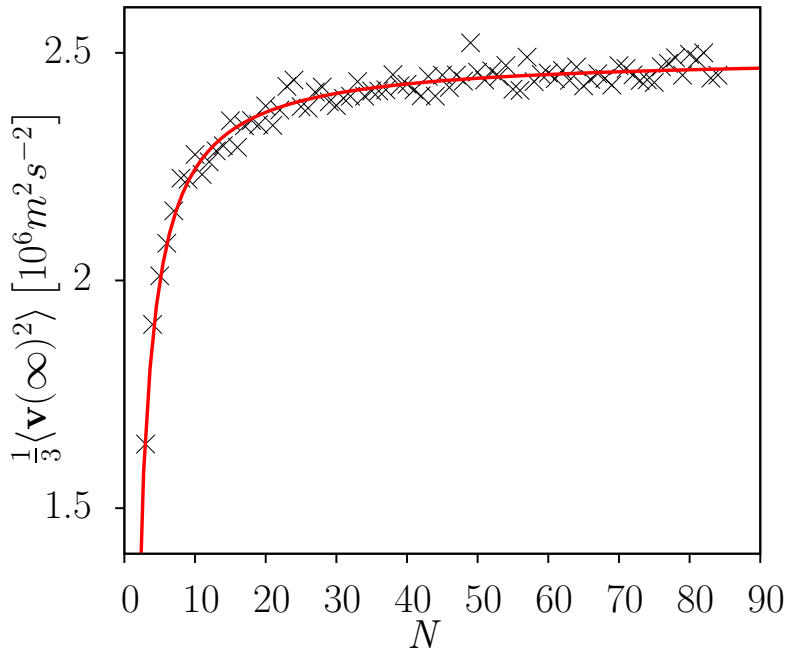


Figure 3.2: Plot of $\frac{1}{3}\left\langle \mathbf{v}(\infty)^2 \right\rangle$ against the number N of non-interacting identical particles. The points are found from 'equilibrated' DL_POLY_4.03 trajectories with parameters $m = 1$ a.u., $\gamma = 100\,\mathrm{ps}^{-1}$, and total simulation time 0.1 ns, and the solid line corresponds to Eq. (**3.5.9**).

The integrals can be evaluated using $\langle \xi_i(t')\xi_i(t'') \rangle = \delta(t' - t'')$ giving,

$$\left\langle v_i^2(t) \right\rangle = v_i(0)^2 \mathrm{e}^{-2\gamma t} + \left(\frac{N-1}{N}\right)^2 \frac{b_i^2}{2\gamma}\left(1 - \mathrm{e}^{-2\gamma t}\right) + \left(\frac{1}{N}\right)^2 \sum_{j=1, j\neq i}^{N} \left(b_j \frac{m_j}{m_i}\right)^2 \frac{1}{2\gamma}\left(1 - \mathrm{e}^{-2\gamma t}\right) \quad (3.5.6)$$

and by taking the limit $t \to \infty$ we find,

$$\left\langle v_i(\infty)^2 \right\rangle = \frac{k_B T}{m_i}\left[1 + \frac{1}{N}\left(\frac{\mathcal{M}}{N m_i} - 2\right)\right], \quad (3.5.7)$$

where $\mathcal{M} = \sum_{i=1}^{N} m_i$ is the total mass of the system. We can then define a temperature of the particle, T_{eff}^i, according to $\frac{1}{2} m_i \left\langle v_i(\infty)^2 \right\rangle = \frac{1}{2} k_B T_{\mathrm{eff}}^i$, such that

$$T_{\mathrm{eff}}^i = T\left[1 + \frac{1}{N}\left(\frac{\mathcal{M}}{N m_i} - 2\right)\right], \quad (3.5.8)$$

which depends on both species mass and the number of particles. The target temperature is returned only in the limit $N \to \infty$. A treatment of the motion in three dimensions gives the same expression for the temperature. The average kinetic energy in three dimensions is then $\langle K \rangle = (1/N)\sum_{i=1}^{N} 3 m_i \left\langle v_i(\infty)^2 \right\rangle / 2 = \frac{3}{2} k_B T \left(1 - \frac{2}{N} + \sum_{i=1}^{N} \mathcal{M}/(N^3 m_i)\right)$.

We now compare this analysis with simulation data. First we study a system of identical non-interacting particles of mass $m = 1$ a.u. in a 3-d non-periodic box. Eq. (**3.5.7**) for the asymptotic

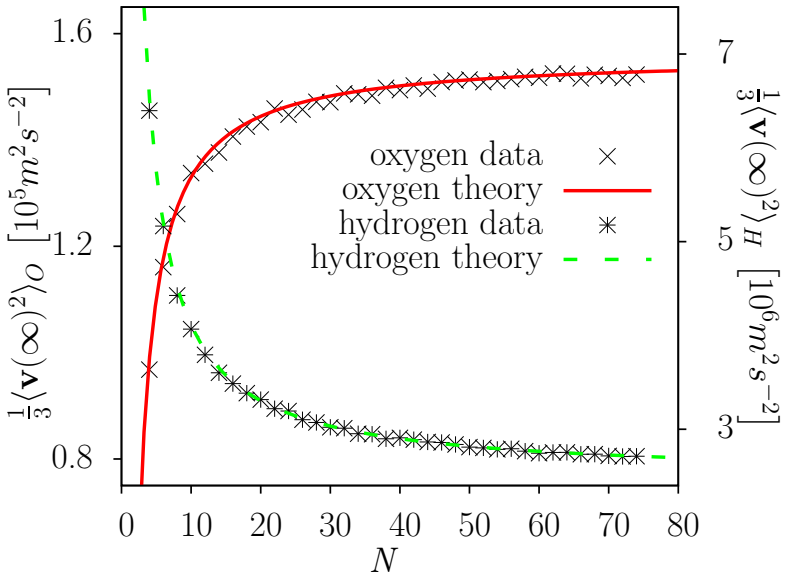


Figure 3.3: Plot of $\frac{1}{3}\left\langle \mathbf{v}(\infty)^2 \right\rangle_\alpha$ against the number of non-interacting particles in a DL_POLY_4.03 simulation, half of which are oxygen and the remaining half are hydrogen. Results for oxygen and hydrogen refer to the left and right hand axes respectively. The points correspond to an average of five 'equilibrated' simulations ($\gamma = 100\,\mathrm{ps}^{-1}$ with a simulation length of 0.1 ns) for each value of N and the curves are defined by Eq. (**3.5.10**).

mean square of one velocity component of the particles leads to,

$$\frac{1}{3}\left\langle \mathbf{v}(\infty)^2 \right\rangle = \left\langle v_x(\infty)^2 \right\rangle = \left\langle v_y(\infty)^2 \right\rangle = \left\langle v_z(\infty)^2 \right\rangle = a\left(1 - \tfrac{1}{N}\right) \tag{3.5.9}$$

where $a = k_B T/m$. Figure **3.2** shows $\frac{1}{3}\left\langle \mathbf{v}(\infty)^2 \right\rangle$ obtained from a set of NVT simulations with target temperature $T = 300\,\mathrm{K}$ such that $a = 2.494\times10^6\,\mathrm{m^2 s^{-2}}$, plotted against Eq. (**3.5.9**), showing that the dependence on N is consistent with the model.

A system containing particles with different masses was chosen for a more detailed testing of Eq. (**3.5.7**). For simplicity the system of N particles was composed of equal numbers of light particles (hydrogen, $m_H = 1$ a.u.) and heavy particles (oxygen, $m_O = 16$ a.u.), again with no mutual interactions. It is again possible to simplify Eq. (**3.5.7**) to give,

$$\frac{1}{3}\left\langle \mathbf{v}(\infty)^2 \right\rangle_\alpha = a_\alpha\left(1 - \frac{\tilde{N}_\alpha}{N}\right) \tag{3.5.10}$$

where $\alpha = H$ or O, $a_\alpha = k_B T/m_\alpha$, $\tilde{N}_H = [3 - (m_O/m_H)]/2 = -13/2$ and $\tilde{N}_O = [3 - (m_H/m_O)]/2 = 47/32$. The results for $N \geq 4$ are summarised in Figure **3.3**. As the system size increases, the average squared velocity component of the heavy atoms increases, while that of the light atoms decreases, in

line with the trends observed in Figure **3.1**.

Returning to our simulations of sulphuric acid, the effective temperatures of the hydrogen, oxygen and sulphur species according to (3.5.8) would be given by $T_{\text{eff}}^{H} = T[1 + 12/N]$, $T_{\text{eff}}^{O} = T[1 - 9/(8N)]$ and $T_{\text{eff}}^{S} = T[1 - 25/(16N)]$ using appropriate atomic masses. This behaviour is consistent with the unexpected thermalisation behaviour illustrated in Figure **3.1**.

3.5.2.2 Gauss' principle of least constraint

Another method for maintaining a condition during the evolution of a system makes use of Gauss' principle of least constraint [**12**]. In contrast to the analysis presented up to now, however, it is a method for the elimination of the total force on the CoM, rather than just the thermal noise force. It may be shown that a constraint whereby the net momentum of the system, $\mathcal{P}$, is conserved

$$\sum_i p_i - \mathcal{P} = 0, \tag{3.5.11}$$

modifies the equations of motion to give

$$m_i \dot{v}_i = -\gamma m_i (v_i - \mathcal{V}) + b_i m_i \left(1 - w_i\right) \xi_i(t) - \sum_{j=1, j\neq i}^{N} b_j m_i w_j \xi_j, \tag{3.5.12}$$

where $\mathcal{V} = \mathcal{P}/\mathcal{M}$ and $w_i = m_i/\mathcal{M}$. As in Equation (3.5.2), we see several noise terms acting on each particle. We set $\mathcal{V} = 0$ and derive a temperature as in Section 3.5.2.1 resulting in

$$T_{\text{eff,G}}^{i} = T\left(1 - \frac{m_i}{\mathcal{M}}\right). \tag{3.5.13}$$

Here G is used to distinguish this from the result in Section **3.5.2.1**. It is interesting to compare Equations (**3.5.13**) and (**3.5.8**). Both approaches thermalise particles to temperatures that depend on system size (or total mass) and species mass. However, the Gauss approach thermalises all species to a temperature below the target, while the approach based on the elimination of thermal noise acting on the CoM typically gives elevated temperatures to the lighter particles while depressing the temperatures of the heavier species. There is one advantage to the Gauss scheme, however, which is that the average kinetic energy of the whole system, with fixed CoM, is reproduced as expected, namely $\langle K \rangle = 3\sum_i k_B T_{\text{eff},G}^{i}/2 = 3(N-1)k_B T/2$ [**119**]. In this respect, eliminating CoM motion

through Gauss' principle of least constraint is a better constraint than one that merely eliminates the total thermal noise.

3.5.3 Discussion

This investigation was motivated by the unintended effect of the Langevin thermostat implemented in DL_POLY_4.03, namely the strange thermalisation of different atomic species to mass and system size-dependent temperatures. This behaviour is a direct consequence of a constraint imposed in the code such that the CoM momentum is not influenced by the noise component of the thermostat. An analysis of the stochastic dynamics introduced by such a scheme leads to mean square velocities and effective temperatures given by Eq. (**3.5.7**) and **(3.5.8)**. It has been shown that the results of simulations with DL_POLY_4.03 are consistent with these expressions.

If an MD scheme is required where the CoM of a system is held fixed, uninfluenced by noise, an appropriate Galilean transformation of the velocity configuration after the implementation of the equations of motion for each timestep might be a workable alternative algorithm. Alternatively Gauss' principle of least constraint allows for fixing of the system's position while keeping the correct average system kinetic energy.

In conclusion it is advisable to check behaviour carefully when using MD codes for systems that are not the usual focus of attention, such as the small molecular clusters that are of interest to us. In version 4.05 of DL_POLY the CoM constraint that produced the strange behaviour we encountered has been removed, such that small systems can be thermalised correctly, in line with the zeroth law of thermodynamics [**120**]. This analysis would suggest that Gauss' principle of least constraint is

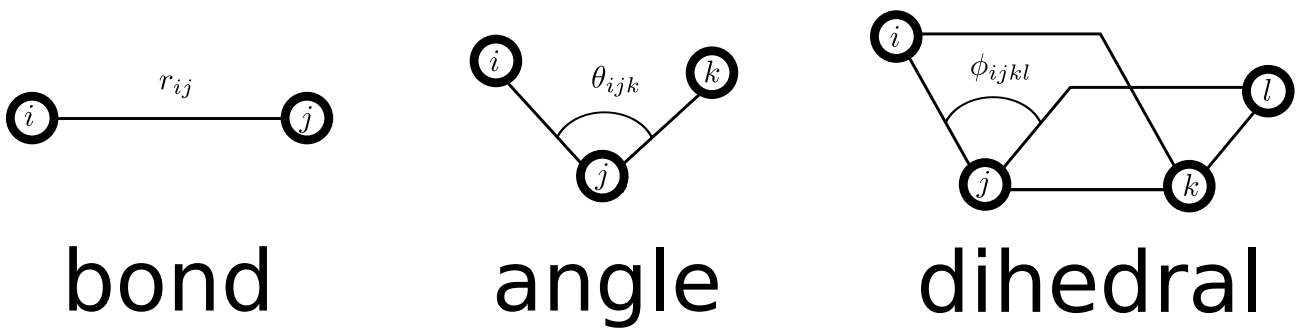


Figure 3.4: The graphical depiction of three typical scalar properties of a molecule that are used for the empirical evaluation of its potential energy. r_{ij} is the distance between the atoms i and j. θ_{ijk} indicates the angle made between atoms i, j and k. ϕ_{ijkl} gives the angle between the two planes defined by atoms i, j, k and j, k, l respectively.

preferred as the expected mean kinetic energy for a CoM constrained system is found. For assurance that the simulations are correctly controlling the temperature of the system after the constraint has been removed the equipartion theorem was tested and the velocity distributions of the atoms were compared against the expected Maxwell-Boltmann distributions. This analysis is described in further detail in Appendix **A**.

3.6 Empirical potential

Molecular dynamics simulations require the evaluation of the potential at each time-step to provide the forces and the system's energy. An empirical potential is typically a function of the particle positions within the system and is simple to implement and computationally inexpensive. This allows a significant number of time-steps to be evaluated. The intra-molecular terms are typically described by a series of bond, angle and dihedral potentials as displayed by Figure **3.4**. The inter-molecular potential, V_{inter}, is commonly a composed of a Coulomb, $V_{Coulomb}$, and a Lennard-Jones, V_{LJ}, potential. One way of evaluating these potentials is as follows,

$$V_{inter} = V_{LJ} + V_{Coulomb} = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] + \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r} \tag{3.6.1}$$

where ϵ and σ are parameters for the Lennard-Jones potential. q_1 and q_2 are the partial charges for the particles and ε_0 is the permittivity of free space. Evaluation of the potential energy of the system is not limited to such empirical functions; an example being the more complex approach of electronic density functional theory which is described in Chapter **4** where the electrons in the system are treated explicitly. These more complex approaches offer several improvements over that which is achievable at the empirical level at the expense of increased computational demand.

3.7 The role of proton transfer

The proton transfer from the sulphuric acid molecule to the water molecule in the sulphuric acid and water system is a major event and has been investigated both via experiment and through theoretical approaches (see Chapter **2**). The ionisation of the cluster is known to increase the nucleation rate of sulphuric acid and water [**10**]. This event is seen as a convenient phenomenon to study in this thesis

as there is a sizeable literature content with which to make comparisons with. Throughout the thesis hydrogen bonds and proton transfers events are used to give analysis on the model.

3.8 Analysis methods

There are several useful technique for analysing trajectories produced from molecular dynamics simulations. Here two such techniques that were used in the project are discussed.

3.8.1 Block analysis

The mean of a given property simulated for N time-steps is given by $\langle A \rangle = (1/N)\sum_{i=1}^{N} A_i$ and the variance is defined as $\sigma^2(A) = \sigma^2 = \langle A^2 \rangle - \langle A \rangle^2$. If each sample is independent then the statistical error in the mean (also known as the standard error) is given by $\sigma^2(\langle A \rangle) = s = (1/N)\sigma^2(A)$ **[108]**. Here σ and s are used to avoid confusion later. However in practice most simulation techniques cannot ensure the independence of each sample. The true error in the mean requires the calculation of the autocorrelation function. Block analysis offers a simpler alternative for estimating the error of a property derived from a molecular dynamics simulation. It is based upon block averaging individual measurements in order to gain an independent sampling. The method involves the use of blocking averages where two adjacent data points are averaged to make one new data point. Therefore each blocking procedure will halve the size of the data set. A key question is then how many of these blocking procedures are required for the data set to represent independent sampling of some property. In the method developed by Flyvbjerg and Petersen **[108, 110, 121]** the error in the mean, s_B, can estimated by,

$$s_B \approx \frac{\sigma_B^2}{N_B - 1} \pm \sqrt{\frac{2\sigma_B^4}{(N_B - 1)^3}} \tag{3.8.1}$$

where N_B represents the number of blocking operations. The subscript B refers to the number of blocking operations that have been performed on the data set. To estimate the number of blocking operations required s_B is plotted against the number of blocking operations performed, B. A plateau region is expected at the centre of the graph where the blocking operation has successfully identified the level needed for independent sampling, but still has enough data points such that the error in s_B is small. Figure **3.5** shows an example produced for a system of one sulphuric acid and three

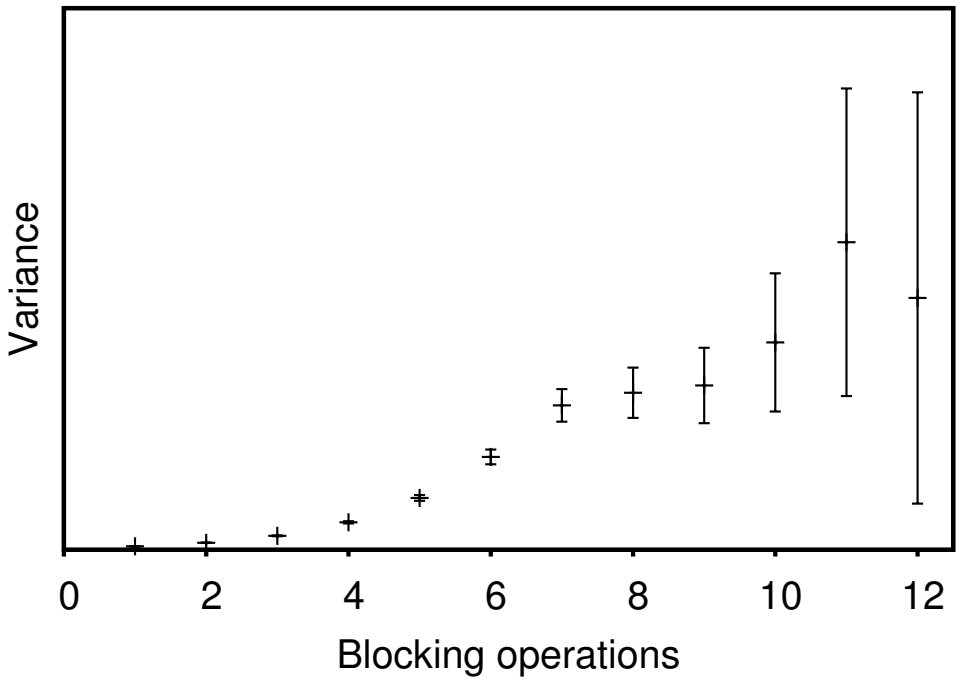


Figure 3.5: The block analysis for the hydrogen bond in trihydrated sulphuric acid system. The system was simulated for 20000 time-steps with each time-step representing 0.5 fs. The points indicate the standard error for the number of blocking operations and the error bars indicate the error in the standard error (see Eq. (**3.8.1**)).

water molecules. The variance in Figure **3.5** is that of a covalent hydrogen-oxygen bond length where the hydrogen is also involved in a hydrogen bond. The plateau region is seen between 7 and 9 blocking operations. The plateau is short which indicates that longer studies are needed to ensure the correct number of blocking operations is identified. However, limitations on computation power have restricted expanding the length. Hence the value of 8 blocking operations or 256 data points was chosen as a reasonable value for independent sampling based on the available data, therefore a blocking length of 0.126 ps is expected to give an independent sample of the desired property. This analysis tool is used in Chapters **4** and **5** to study clusters of sulphuric acid and water.

3.8.2 Potential of mean force

The potential of mean force, W, is defined as the amount of work required to remove a particle from its current location to a position at infinity whilst the remainder of the system is allowed to relax according to the appropriate equilibrium ensemble statistics. A two dimensional example of a

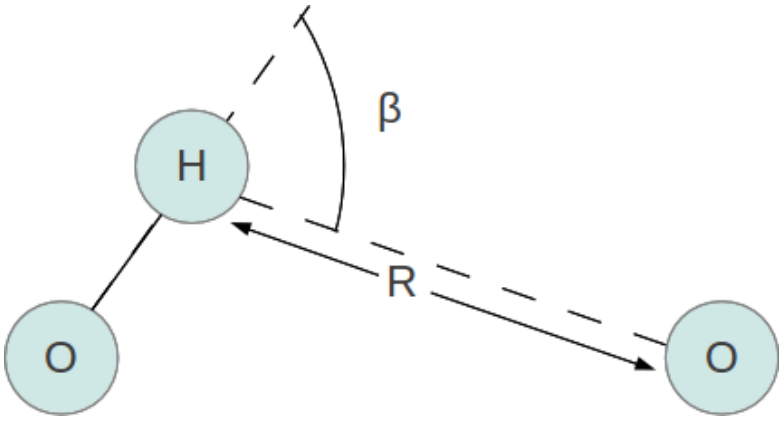


Figure 3.6: Depiction of parameters R and β for the definition of a potential of mean force.

potential of mean force can be related to the distribution given by a simulation via [**122**, **123**],

$$W(R,\beta) = -kT \ln g(R,\beta) \tag{3.8.2}$$

where $g(R,\beta)$ is defined as the ratio of the number of data points observed in a given shell (defined by the R and β) to the expected number of data points if the particles were non interacting. R and β are used to provide a two dimensional description of the hydrogen bond in the system. There are various parameters which can be used to study the hydrogen bond [**122**]. It was decided to use the definitions as outlined in Figure **3.6**. Once the parameters are defined the denominator can be determined, using R and β the denominator is given by $2\pi\rho\sin(\beta)R^2 d\beta dR$ with ρ being the molecular density of the system.

The potential of mean force provides a useful tool for studying the potential energy landscape of a configuration where there is a large number of degrees of freedom. Specifically, the stability of configuration can be studied by looking at the depth of the free energy well that is produced from the potential of mean force. An interesting PMF figure is investigated in Figure **4.7b**.

Chapter 4

Density Functional Theory

One of the most popular quantum chemistry methods in recent years has been that of electronic density functional theory, DFT [**53**]. In this chapter the theoretical framework for the method is presented. The chapter continues by employing the DFT method to provide a potential energy surface for molecular dynamics simulations of small clusters of sulphuric acid and water. The simulations are then analysed and conclusions are drawn about configuration and dynamic properties of these small systems. The work serves as a preliminary study before the zero-points effects on these structures are considered in Chapter **5**.

4.1 Electronic density functional theory

4.1.1 Background

A short history of DFT is given by Perdew and Ruzsinszky [**53**] its popularity in chemistry came many years after the theory was first presented as the second generation of functionals gave much improved descriptions of atomistic energies relevant to chemistry applications. The book presented by Martin [**47**] reviews the theory of DFT and a useful summary of issues concerning DFT is provided by Perdew et al. [**124**]. Payne *et al.* [**125**] gives a practical guide for implementing DFT for MD simulations. The modern version of DFT was introduced by two seminal papers in the mid 1960s. The first by Hohenberg and Kohn [**51**] in 1964 introduced the following two theorems [**47**]:

Theorem I The ground state particle density, $n_0(\mathbf{r})$ can uniquely determine, except for a constant, the external potential $V_{ext}(\mathbf{r})$ acting on any system of interacting particles.

Theorem II A *universal functional* for the energy, $E[n]$ in terms of the ground state particle density

$n_0(\mathbf{r})$ can be found which is valid for any external potential. The density, $n(\mathbf{r})$ which minimises $E[n]$ is the exact ground state particle density, $n_0(\mathbf{r})$ for $V_{ext}(\mathbf{r})$.

These theorems are known as Hohenberg-Kohn theorems **[47]**, simple proofs exist for both theorems (see Appendix **B**). Using a functional of the density over the usual wavefunction approach vastly reduces the parameter space of the system being studied. However these theorems tell us little about practicalities of performing electronic structure calculations using DFT. In addition the scaling of system size or length of simulation causes computation demands that limits the scope of the method. Reference **[126]** gives an overview of DFT and a discussion of the scaling problem in many body systems.

In 1965 Kohn and Sham **[52]** proposed an ansatz that aimed to solve the many body problem via the creation of an auxiliary system. The auxiliary system reduces the problem to a one particle system with an external potential introducing many body effects. The auxiliary system representation ansatz relies on two assumptions **[47]**:

Assumption I The exact ground state density can be represented by the ground state density of an auxiliary system of non-interacting particles. This assumption is known as "non-interacting-V-representability".

Assumption II The auxiliary Hamiltonian is chosen to have the same kinetic operator as a single particle system. An effective *local* potential operator is introduced $V^{\sigma}_{eff}(\mathbf{r})$ acting on an electron of spin σ at position $\mathbf{r}$. The local form of the operator is not essential, however, it is useful as it simplifies the work required to find the ground state density.

With these two assumptions the many body system has been simplified by removing electron-electron interactions to make them non-interacting. These interactions cannot be ignored without an accuracy penalty and so are included in the effective local potential. The exact form of this effective local potential is not known, but simple approximations have been found to give somewhat surprisingly satisfactory results in certain systems. The first attempt to approximate $V^{\sigma}_{eff}(\mathbf{r})$ is known as the local density approximation (LDA) introduced in the seminal paper by Kohn and Sham **[52]** in 1965. DFT has been developed for different uses under schemes such as the generalised gradient approximation and hybrid functionals.

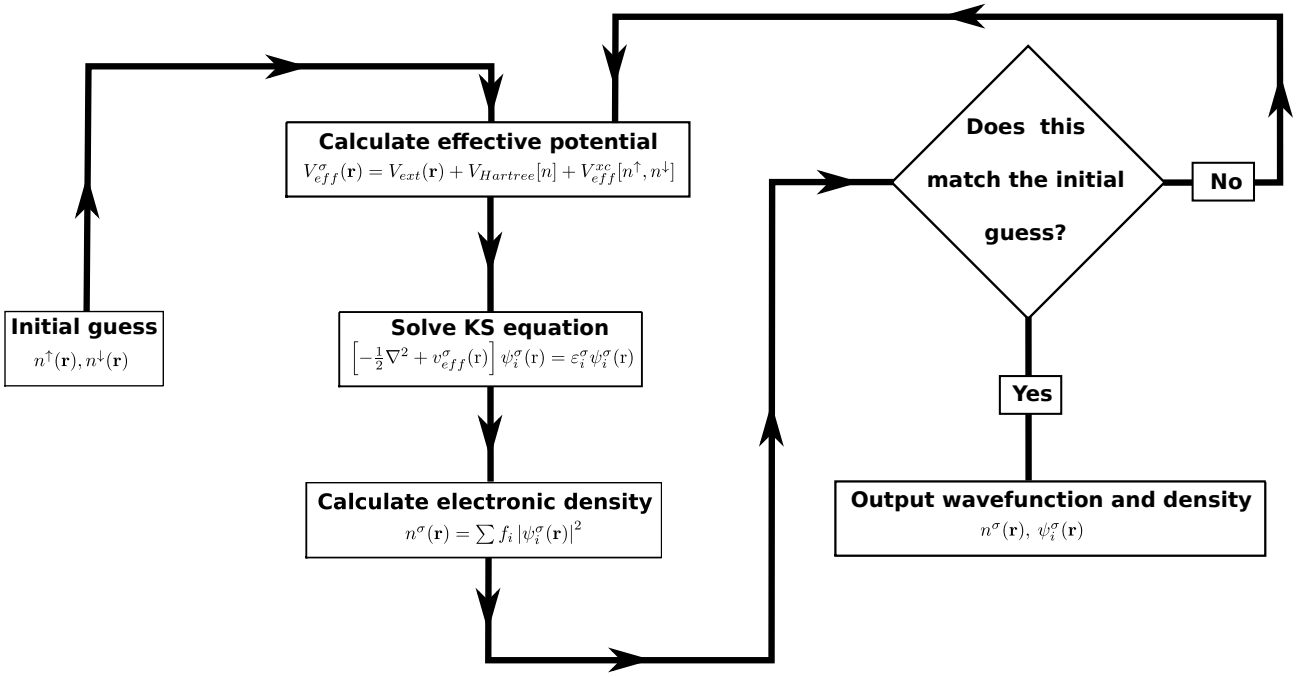


Figure 4.1: Schematic of the self consistent field (SCF) iterations used in many DFT programs to find the ground state density and corresponding wavefunction . The image is equivalent to Figure **9.1** on page 173 in reference **[47]**.

This project used the CASTEP 5.5 **[127]** code which is one of many available software packages which implements DFT. Alternatives to DFT include Hartree-Fock, Møller-Plesset perturbation theory**[48]** and Coupled Cluster **[128]** methods.

The standard implementation of DFT is that of a self consistent iterative procedure. This method uses the connection between the wavefunction of the system and the density of the electrons in a system, i.e. $n(\mathbf{r}) = \sum_{\sigma}\sum_i \left|\psi^{\sigma}_i(\mathbf{r})\right|^2$ where i labels the orbitals and it is summed over the spin states, σ. This wavefunction is a solution of the single particle Schrödinger-like equation with the auxiliary Hamiltonian. Figure **4.1** shows the schematic given on page 173 of the *Electronic Structure* book by Martin**[47]** and describes the general self consistent scheme used in many implementations of DFT.

The one particle like Schrödinger equation reduces the amount of work required by the DFT method to calculate the ground state density of the system but is reliant upon the accuracy of the total energy functional whose exact form is unknown. In particular the exchange and correlation contributions[1] are not known from first principles, the functional that quantifies their contribution is referred to as $E_{XC}[n(\mathbf{r})]$. Not all is lost. Simple approximations used for the form of the functional have proved successful **[53]**. The Local Density Approximation, LDA, was the first approximation

[1]The exchange energy arises from the freedom to change labels within the simulation. The correlation energy is due to the repulsive Coulomb interaction between electrons **[124]**.

[52]. It assumes that the exchange and correlation energy per electron at every point is equal to that of homogeneous electron gas (jellium) with the density equal to that of the density at that point in space,

$$E_{XC}[n(\mathbf{r})] = \int \varepsilon_{XC}(n(\mathbf{r}))n(\mathbf{r})d\mathbf{r}$$

Here $\varepsilon_{XC}(n(\mathbf{r}))$ is the exchange correlation energy per particle as a function of the density. We can then construct the total energy of the Kohn-Sham equation, E_{KS}, as,

$$E_{KS}[n(\mathbf{r})] = T[n(\mathbf{r})] + \int d\mathbf{r} V_{ext}(\mathbf{r})n(\mathbf{r}) + \int d\mathbf{r}d\mathbf{r}' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{II} + E_{XC}[n(\mathbf{r})]$$

where $T[n(\mathbf{r})]$ is the kinetic energy and E_{II} is the nuclei-nuclei potential energy.

4.1.2 Generalised Gradient Approximation

LDA was found to work well with bulk surfaces and their surfaces **[53]**. The generalised gradient approximation, GGA, is the next generation of functional and improved the molecular energies making the method more applicable for chemistry applications **[53]**. Here it is the assumed that ε_{XC} is dependent both on the density and the gradient of the density, i.e. $E_{XC}[n(\mathbf{r}), \nabla n(\mathbf{r})]$. This assumption does not mean that the GGA is an improvement on LDA in all cases, this is clear because we do not know the exact form of E_{XC}. Another class of functionals called the hybrid functionals mix LDA, GGA and pure Hartree-Fock exchange **[47]** energies, a popular example of this type functional is the B3LYP functional **[61]**.

Hydrogen bonds are of particular interest in this project because they are weak flexible bonds and are important structurally for liquid systems. Due to their comparative weakness they have typically been difficult to simulate to a high accuracy. Proton transfer within the water-hydronium system is known as the Grotthuss mechanism and the light weight of the proton is thought to be the reason for the unusually high diffusion rate of an excess proton in a water network **[80]**. For this study the Perdew-Burke-Ernzerhof (PBE) functional **[79]** was chosen due to its performance with hydrogen bonds being of good quality **[129–131]** and its availability in the CASTEP program.

4.1.3 Bloch theorem

The Bloch theorem [**47**, **132**, **133**] states that the wavefunction of an particle in periodic environment such as an electron in a crystal can be expressed as a plane wave ($e^{i\mathbf{k}\cdot\mathbf{r}}$) multiplied by a periodic function ($u_{\mathbf{k}}(\mathbf{r})$), which has the periodicity of the periodic environment),

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}) \tag{4.1.1}$$

where $\mathbf{r}$ is the position vector, $\mathbf{k}$ is the wave vector and has units of inverse length. The momentum of the state defined by $\mathbf{k}$ is given by $\hbar|\mathbf{k}|$.

4.1.4 Plane wave basis set

In the context of quantum chemistry, a basis set consists of a collection of functions. The wavefunction is then constructed as a superposition of basis set functions, the wavefunction must obey the relevant governing equation for the system.

The independent electron Schrödinger-like equation is the relevant governing equation for DFT and has the form,

$$\left[-\frac{\hbar^2}{2m_e}\nabla^2 + V_{eff}(\mathbf{r})\right]\psi_j(\mathbf{r}) = \hat{\mathcal{H}}_{eff}\psi_j(\mathbf{r}) = \varepsilon_j\psi_j(\mathbf{r}) \tag{4.1.2}$$

here m_e is the electron mass. It is convenient to expect the wavefunction, $\psi_j(\mathbf{r})$ to be normalised (i.e. $\int_{all\ space} |\psi_j(\mathbf{r})|^2\, d\mathbf{r}$) and to observe periodic boundary conditions. Expanding the eigenfunction solution to Eq. (**4.1.2**) as a complete set of Fourier components[2] $\psi_j(\mathbf{r})$ can be written as,

$$\psi_j(\mathbf{r}) = \sum_{\mathbf{q}} c_{j,\mathbf{q}} \frac{1}{\sqrt{\Omega_{cell}}} e^{i\mathbf{q}\cdot\mathbf{r}} \equiv \sum_{\mathbf{q}} c_{j,\mathbf{q}} \,|\mathbf{q}\rangle \tag{4.1.3}$$

where Ω_{cell} is the volume of the cell and $\mathbf{q}$ represents the wave vector of the plane wave which in general points in the same direction as the propagation of the plane wave. In this form we see that the wavefunction has been expanded in a basis set of orthogonal plane waves with coefficients $c_{j,\mathbf{q}}$. Using a plane wave basis set has the benefits of being in a convenient format for the fast Fourier transform (FFT) procedure and can significantly improve the computational performance of an implementation

[2] It is valid to expand any periodic function as a complete set of Fourier components. However for practical reasons only a finite expansion is performed, this is discussed further in 4.1.5

of DFT.

4.1.5 k point sampling and cut-off energy

A crucial part of DFT is the calculation of electronic densities in the system. For this purpose it is essential to perform an integration over the **k** points within the Brillouin zone[3] [**47**]. This sampling can be performed using methods such as that proposed by Chadi and Cohen [**134**] and Monkhorst and Pack [**135**] which replace the integral with a finite sum. In the case of this study further simplifications can be made by observing that we are in fact interested in an isolated cluster and therefore the choice of one **k** point at $(0,0,0)$ is a sensible choice and is known as the Gamma point (Γ).

We have not yet discussed the periodic part of the Bloch theorem solution from Eq. (**4.1.1**). The first step is to substitute the Bloch theorem solution into the single particle like Schrödinger equation. This can be reduced to the form,

$$\sum_{\mathbf{G}'}\left[\frac{\hbar^2}{2m_e}|\mathbf{k}+\mathbf{G}|^2\delta_{\mathbf{G},\mathbf{G}'}+V_{eff}(\mathbf{G}-\mathbf{G}')\right]c_{i,\mathbf{k}+\mathbf{G}'}=\varepsilon_i c_{i,\mathbf{k}+\mathbf{G}'} \tag{4.1.4}$$

$$V_{eff}(\mathbf{G})=\frac{1}{\Omega_{cell}}\int_{\Omega_{cell}}V_{eff}(\mathbf{r})e^{-i\mathbf{G}\cdot\mathbf{r}}d\mathbf{r} \tag{4.1.5}$$

here $\delta_{\mathbf{G},\mathbf{G}'}$ is the Kronecker delta function which is zero for $\mathbf{G}\neq\mathbf{G}'$ and unity for $\mathbf{G}=\mathbf{G}'$. In principle the sum over $\mathbf{G}'$ should be infinite. Fortunately the coefficients of the plane waves $c_{i,\mathbf{k}+\mathbf{G}'}$ typically become small[4] for large $|\mathbf{G}|^2$ as this is related to the kinetic energy of the state. This means we can achieve a good approximation with a finite sum over $\mathbf{G}'$. Many available codes performs this by imposing the following restriction (this is known as the cut-off energy) on the kinetic energy part of Eq. (**4.1.4**) such that,

$$\frac{\hbar^2}{2m_e}|\mathbf{k}+\mathbf{G}|^2<E_{cut}$$

Unfortunately the core electrons are problematic to this limit as their proximity to the atomic nuclei gives a rapidly changing potential energy landscape resulting in a large kinetic energy states. However, the core electrons are usually of little interest in simulations as they do not take part in bonding and are generally of little consequence to reactions. The use of a pseudopotential can

[3]The Brillouin zone is a uniquely defined primitive cell in reciprocal space. For more information see reference [**47**].
[4]The vanishing of the $c_{i,\mathbf{k}+\mathbf{G}'}$ depends upon the form of the potential.

alleviate some of the issues caused by the core electrons. A pseudopotential replaces the real potential with one which has a much softer gradient within a certain distance of the nuclear cores. Pseudopotentials are discussed further in the next section.

4.1.6 Born-Oppenheimer approximation

The Born-Oppenheimer approximation is commonly used to separate the nuclear and electronic wavefunctions to simplify quantum chemistry calculations. The first step is to rewrite the system wavefunction as a product of a nuclear wavefunction and an electronic wavefunction which depends upon the nuclear positions such as,

$$\Psi(\{\mathbf{r},\mathbf{R}\}) = \sum_i \chi_i(\{\mathbf{R}\}) \times \phi_i(\{\mathbf{r}\};\{\mathbf{R}\}) \tag{4.1.6}$$

it is then possible to write two Schrödinger equations for the electronic and nuclei wavefunctions which need to solved simultaneously. It is complicated by considering all crossing terms between these equations which describe the effects of a changing nuclear configuration on the electrons. In effect the Born-Oppenheimer approximation states that the electronic state of the system responds instantaneously in respect to the nuclear positions. However the electrons are allowed to gain kinetic energy through translation motion matching the nuclei. Another way of stating this is that the electronic state is always relaxed with respect to the motion of the nuclei. The Born-Oppenheimer approximation can also be inferred from the observation of the difference in the average velocity of the two particles. Typically the nuclei move slower than the electrons by a factor of about 10^3 **[133]**. This is usually combined with the neglect of the translation kinetic energy gained by the electrons which allows us to define separated nuclear and electronic equations **[47]**,

$$\left[\hat{T}_e(\{\mathbf{r}\}) + \hat{V}_{eN}(\{\mathbf{r}\};\{\mathbf{R}\}) + \hat{V}_{ee}(\{\mathbf{r}\})\right] \phi_i(\{\mathbf{r}\};\{\{R\}) = E_e\phi_i(\{\mathbf{r}\};\{\mathbf{R}\}) \tag{4.1.7}$$

$$\left[\hat{T}_N(\{\mathbf{R}\}) + \hat{V}_{NN}(\{\mathbf{R}\}) + E_e(\{\mathbf{R}\})\right] \chi_{si}(\{\mathbf{R}\}) = E_N\chi_{si}(\{\mathbf{R}\}) \tag{4.1.8}$$

The Born-Oppenheimer approximation is responsible for the classical treatment of the nuclei which is permitted given the separation of the electronic and nuclei wavefunction and is justified by the difference in the mass. This assumption of classically behaved nuclei is examined in Chapter **5**.

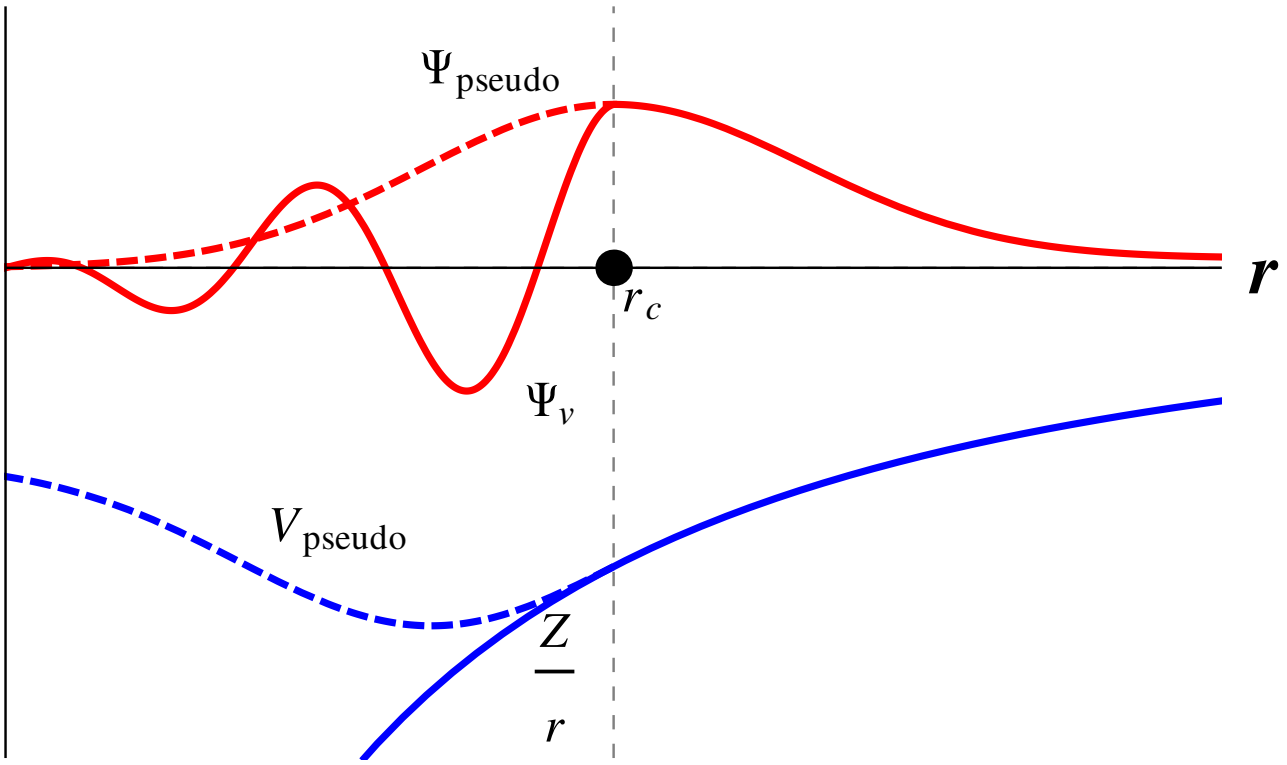


Figure 4.2: Imaging depicting the Coulomb potential (blue line), pseudopotential (blue dashed), all electron wavefunction (red line) and pseudopotential wavefunction (red dashed line). This image is a recreation of Figure 5 in reference [**125**].

4.1.7 Pseudopotential

The pseudopotential is a method used for representing the potential felt by an electron near the atomic nuclei. The Born-Oppenheimer approximation as described in Section **4.1.6** separates the electronic and nuclei wavefunction, and the nuclei are assumed to behave as classical point-like particles. The potential felt by the electrons near the atomic nuclei is steep due to the strong Coulomb potential. This causes problems with the cut off energy convergence as discussed in the previous section. A solution to this problem is to replace the potential for the valence electrons with a pseudopotential within a given radius of the atomic core. The effect is a smoother potential gradient leading to lower kinetic energies near the core and a smaller basis set. The situation is justified by the argument that electrons occupying a closed shell within an atom are rarely important for chemistry. This idea is extended by not considering core electrons, i.e. the oxygen electronic configuration $1S^2 2S^2 2P^4$ is represented by the six electrons in the second angular momentum shell [He]$2S^2 2P^4$. The electrons in the [He] shell can be treated as shielding the positively charged nucleus. The pseudopotential can incorporate the core electrons efficiently into the potential experienced by the valence electrons

lowering the number of electrons which are treated explicitly. The pseudopotential is usually fitted such that it reproduces the scattering properties of the valence electrons in isolated atoms [**125**].

There are two widely used types of pseudopotential which have been developed. They are known as norm conserving and ultrasoft. The difference is due to the treatment of the normalization condition i.e. satisfying $\int |\Psi(\mathbf{r})|^2 d\mathbf{r} = 1$. This condition is imposed in the norm conserving type and is not imposed in the ultrasoft type[5]. In this work the automatically generated ultrasoft pseudopotentials from the CASTEP program were used as there was insufficient time in the project to assess the use of different pseudopotentials. Figure **4.2** shows the general shape of a wavefunction and potential as a function of the distance from a nuclei core. The solid line represents the all electron picture and the dashed line represents the pseudopotential and corresponding wavefunction. The pseudopotential matches exactly the all electron potential for distances larger than or equal to the cut off distance, r_c. The cut off distance r_c is the distance at which the form of the potential felt by the valence electrons changes from the all electron potential to the pseudopotential. A more detailed review of pseudopotentials can be found in reference [**136**].

4.2 Simulation details

This section details the choices made for the DFT model used to perform simulations in this project. The main goal of the study was to determine the importance of zero point motion in small clusters of sulphuric acid and water which involved comparing the results of standard DFT against path integral molecular dynamics (see Chapter **5**). However, some DFT level results were also obtained and are discussed in Section **4.3**.

Molecular dynamics simulations were run at 300 K incorporating both classical nuclear dynamics and path integral molecular dynamics (see Chapter **5**) using the CASTEP [**127**] (version 5.5) code. The standard on-the-fly ultrasoft pseudopotential provided internally by the CASTEP code was employed for all calculations. The PBE functional was used with a plane wave basis set. The PBE functional [**79**] (see Section **4.1.2**) has been found to perform well for hydrogen bonded systems[**130**, **131**]. A cut off energy of 550 eV was found to be a suitable choice for convergence as shown in Figure **4.3**. A larger cut off energy would require more computational time reducing the scope of the study. A

[5]The ultrasoft pseudopotential can typically choose a larger cut off radius (r_c in Figure **4.2**) which can increase the smoothness of the potential reducing the need for a large cut off energy.

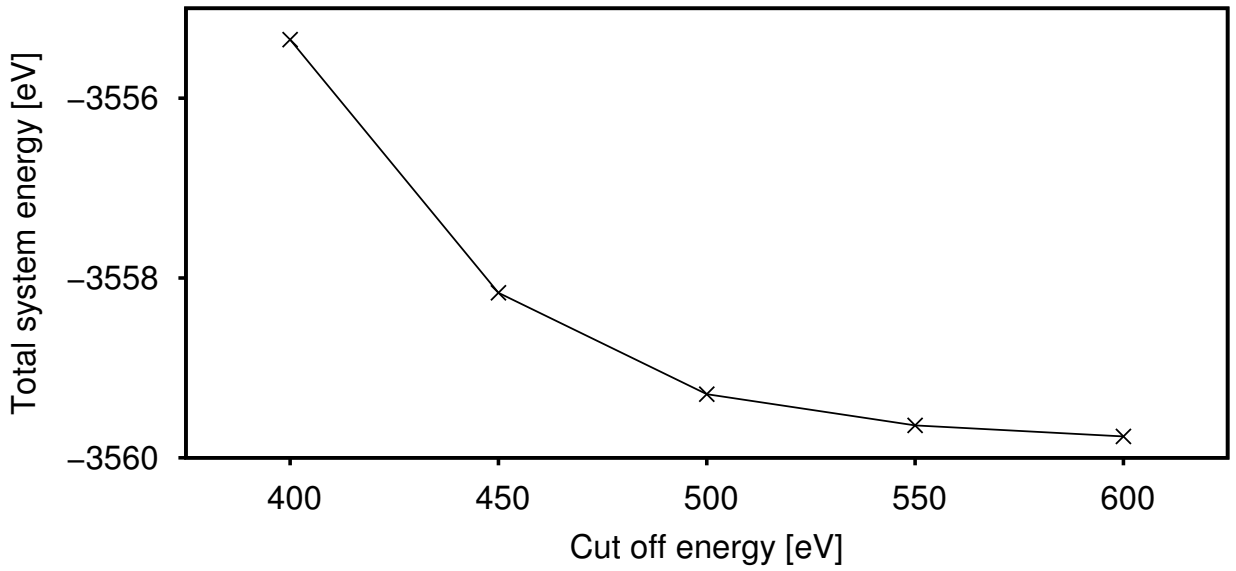


Figure 4.3: Comparison of cut off energy for the plane wave basis set against total energy of a geometry optimised trihydrated sulphuric acid configuration as shown in Figure **4.5a**.

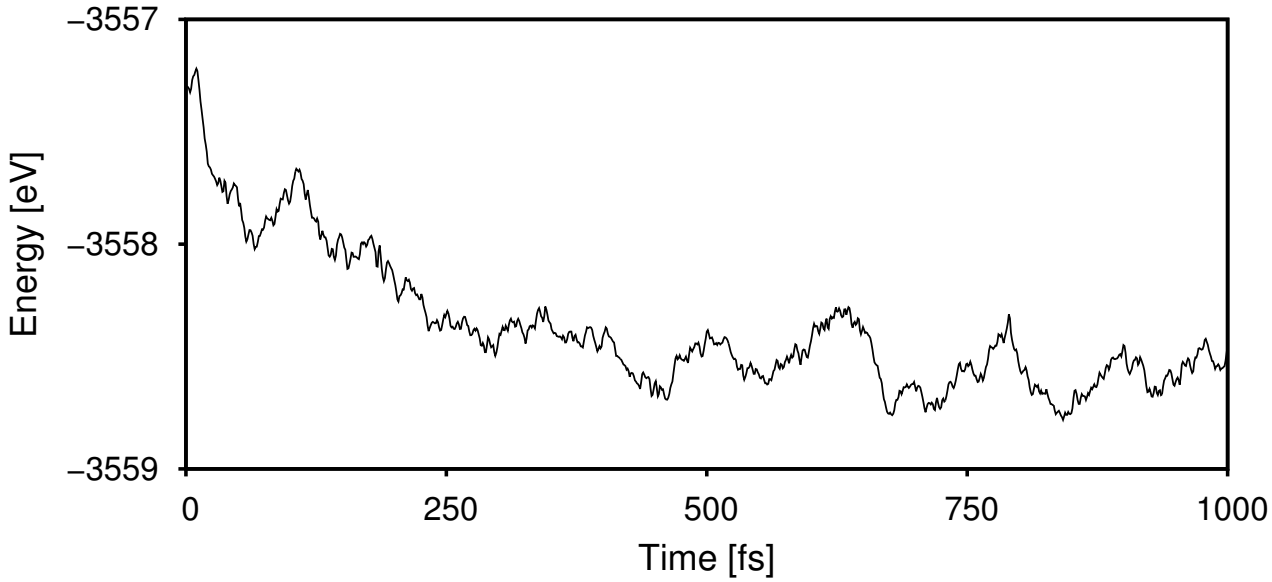


Figure 4.4: Plot of the system's potential energy as a function of time for the first picosecond of the simulation referred to as SATH 1 bead in Table **5.1**. The plot illustrates the typical equilibration period as seen by the potential energy of the system. Statistics were collecting in order to perform analysis once the system had reached a steady state, in this case after half a picosecond.

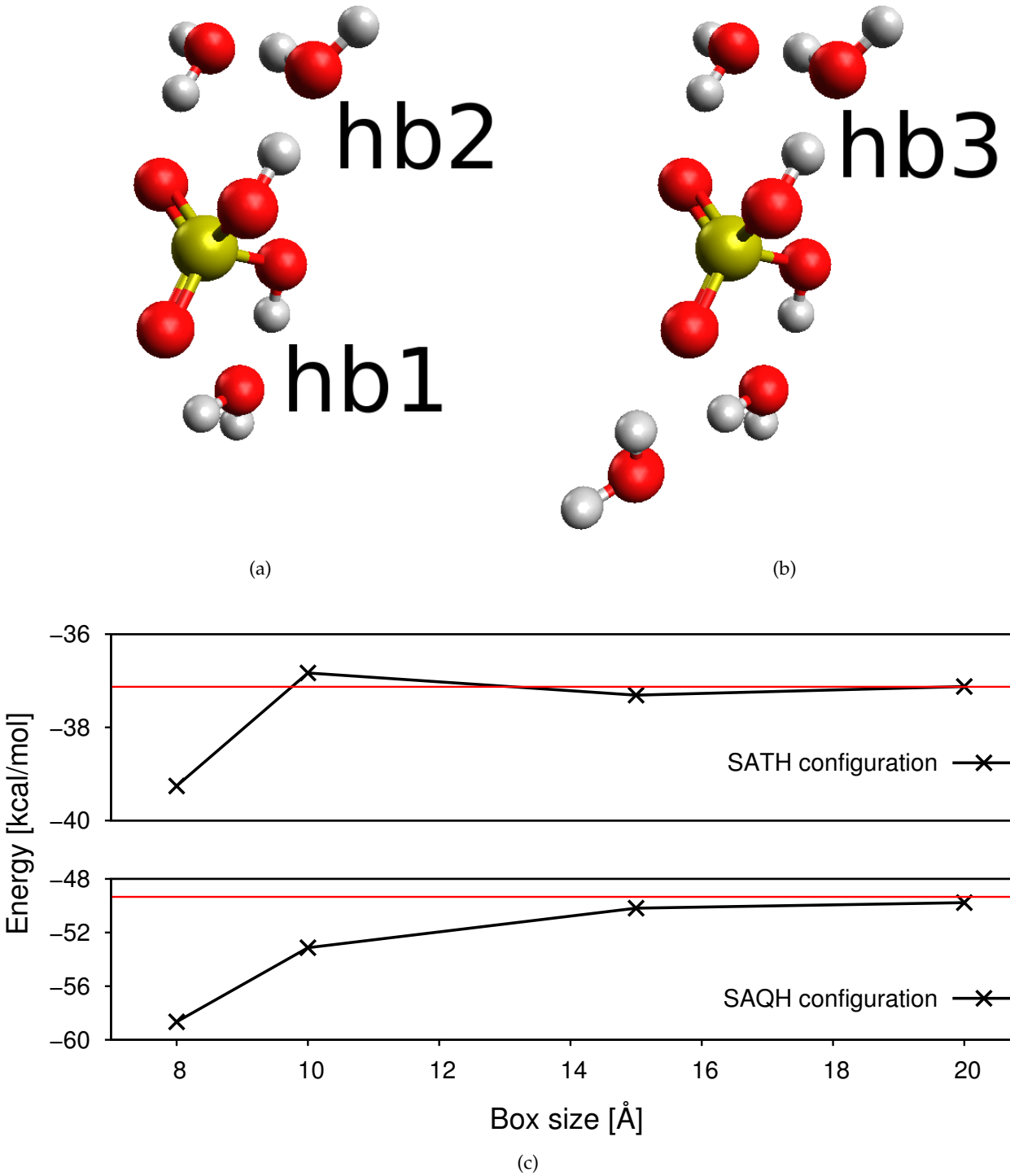
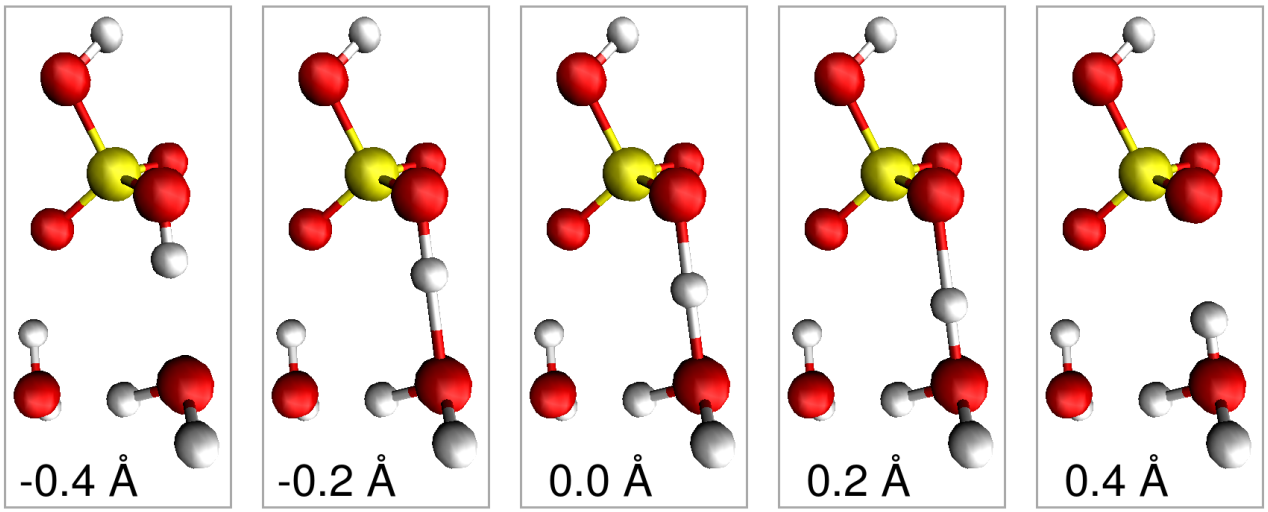
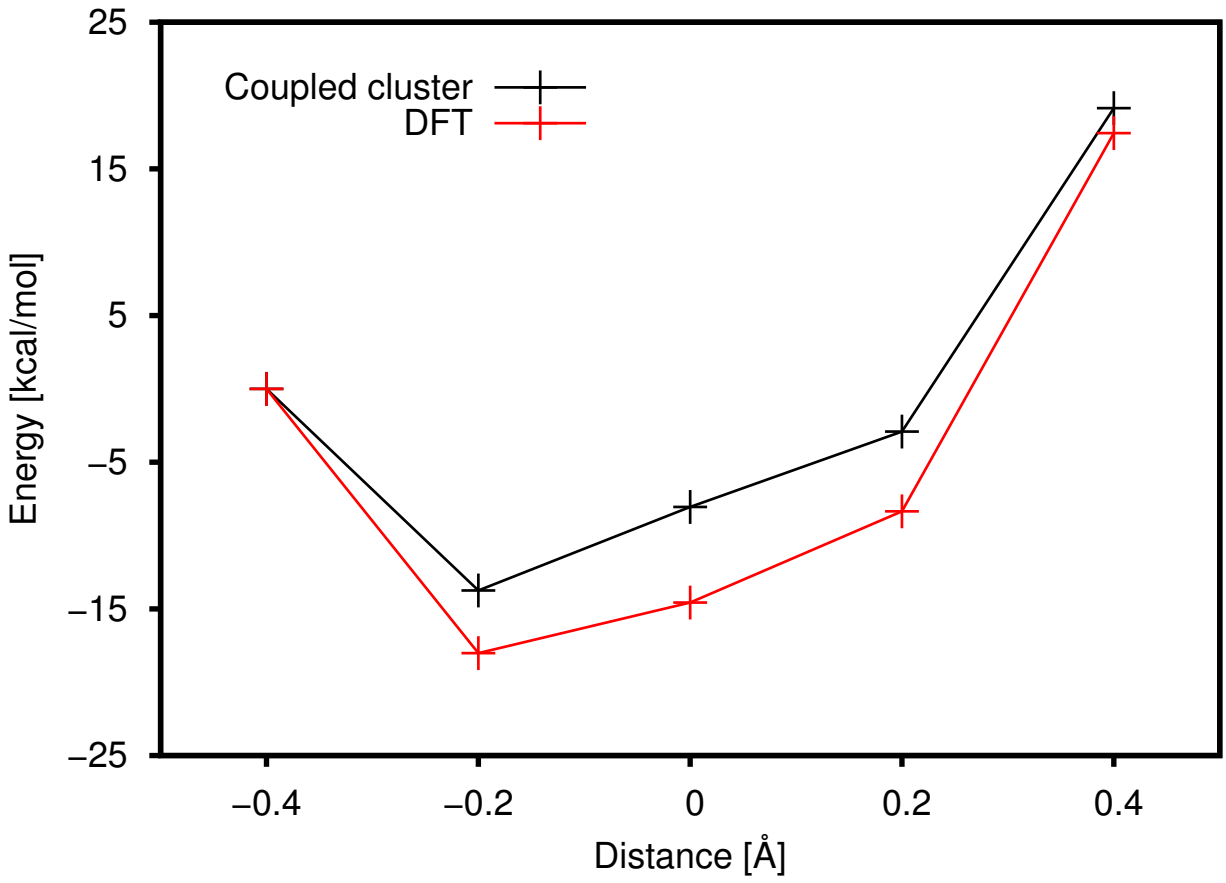


Figure 4.5: Geometry optimised configurations for tri- and tetrahydrated sulphuric acid (SATH and SAQH) clusters are shown in (A) and (B) respectively. (C) shows the binding energies of configurations (A) and (B) as a function of the system box size obtained at the DFT level converging to values obtained at the MP2 level by Temelso *et al.* **[69]**. It is important to note that the DFT simulations did not use the same set of coordinates as the MP2 simulation, rather configurations (A) and (B) were recreated and then a geometry optimisation was run to find the total energy of the configuration.



(a)



(b)

Figure 4.6: Image (B) shows the energy of a sulphuric acid dihydrate system as a function of proton position as depicted in image (A). The system was first geometry optimised under the DFT scheme as presented here. The proton was then moved along the oxygen(sulphuric acid)-oxygen(water) vector in steps of 0.2 Å from the centre of the oxygen-oxygen vector with the negative direction going towards the sulphuric acid (as depicted in (A)). For each point a separate coupled cluster calculation was run with an identical configuration.

	time-step [fs]	Simulation time [ps]
SATH	1.00	11.000
SAQH	1.00	1.000
config H	1.00	10.900

Table 4.1: Compilation of the simulation length and time-step for the MD runs performed. SATH refers to sulphuric acid trihydrate and SAQH refers to sulphuric acid tetrahydrate and correspond to structures shown in Figures **4.5a** and **4.5b** respectively. Config H refers to the trihydrate configuration shown in Figure **4.7a**.

time-step of 1 fs was used for all standard DFT simulations[6]. CASTEP utilizes the Born-Oppenheimer version of ab initio MD which performs the standard DFT SCF iterations at each time-step. This procedure benefits from using the previous time-step's wavefunction as a starting guess of the current time-step's wavefunction. The Langevin thermostat with a friction constant of 0.01 fs^{-1} was found to control the temperature satisfactory (see A.1) and was used for all DFT simulations. The equilibration period was judged by observing when the running mean of the system's potential energy of the system had relaxed (usually requiring 0.5 ps) and also by monitoring the distribution of cluster 'temperature' (or kinetic energy in the centre of mass frame), which ought to be approximately Gaussian [**109**] with a standard deviation (σ) obeying $\sigma / \langle T \rangle \sim N^{-1/2}$ (see Appendix **A**). Figure **4.4** shows a typical relaxation period of the system's energy during the initial phase of the molecular simulation. Initial configurations of sulphuric acid and water identified from the literature [**58–60**] were constructed under a classical potential (MMFF94s) using the Avogadro [**137**] (version 1.0.3) package. The choices of time-step and simulation time for various cases are given in Table **4.1**.

The box size of the system was optimised against MP2 level data [**69**] as shown in Figure **4.5c**. The binding energies, at zero temperature, of the two configurations in Figures **4.5a** and **4.5b** are compared against MP2 level data. A box size of 15 Å was chosen as a compromise between accuracy and computational demand. The NERSC supercomputer was used to run all DFT MD simulations, a typical job involved 48 processors for 10 hours for 1000 time-steps per bead.

One way to check if the simulations are running as expected is to test the calculations obtained here at the DFT level of theory against higher level theory such as MP2 or CC. It was possible to compare against CC calculations thanks to Dr. Theo Kurtén. Due to the computational overheads of running CC calculations we decided to focus our effort on proton transfer between the sulphuric acid and a water molecule. Five calculations were ran where the geometry of a sulphuric acid dihydrate

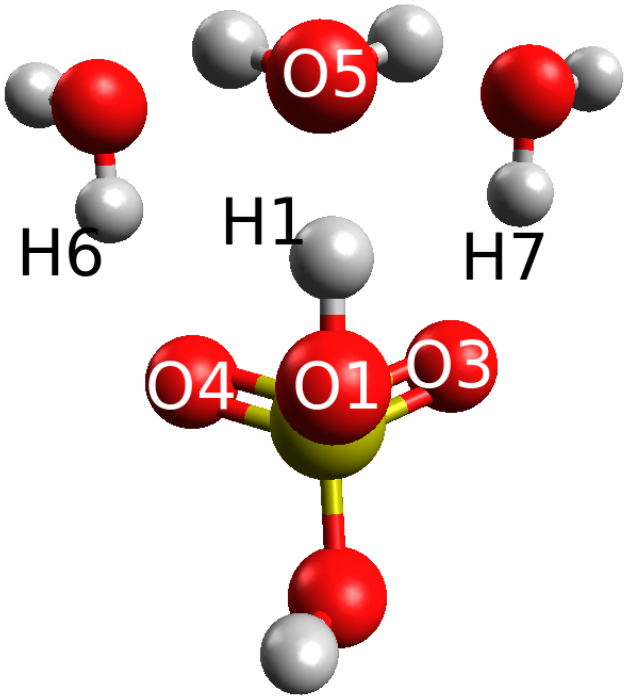
[6] A simulation ran with a 2 fs time-step was found to give an unstable equilibration period in comparision with the 1 fs time-step simulations.

was optimised under the DFT method under the CASTEP code as described here. The geometry optimised structure in the position of one hydrogen which was moved along an oxygen-oxygen vector in steps of 0.2 Å as shown in Figure **4.6a**, the negative direction is towards the sulphuric acid and the positive direction is towards the water molecule. The Molpro code [**138**] was used to run CC calculations with single and double excitations. Figure **4.6b** shows the energy of the system over the distance moved by the single proton. Although this is not a thorough test of the calculations performed here it is reassuring to see a similar energy landscape for the transferring proton between the two theories. A timescale of a month was required to perform the CC calculations whereas the DFT calculations could be done within hours demonstrating that is not feasibility to perform CC MD with current technology.

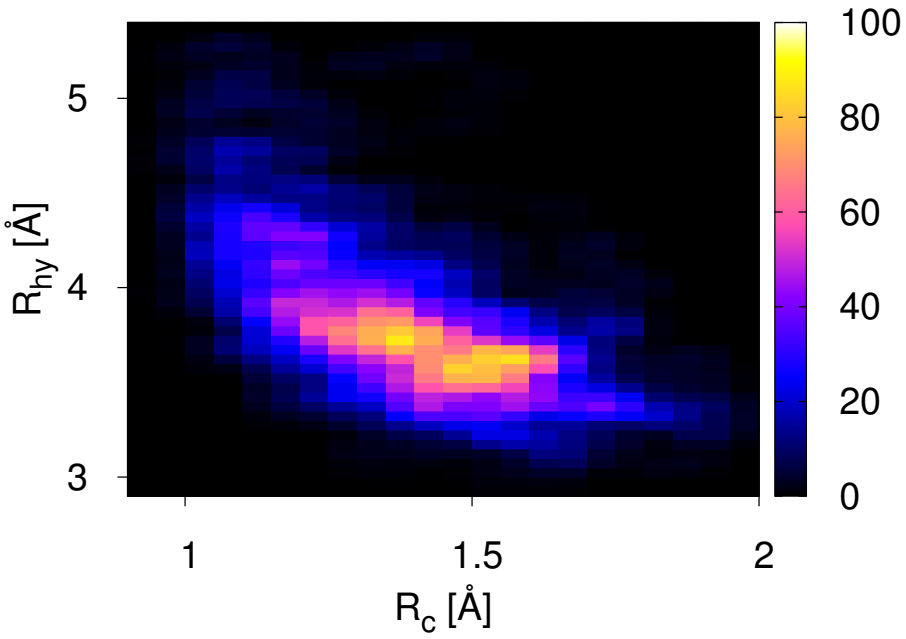
4.3 DFT results

Molecular configurations likely to feature a dissociated sulphuric acid molecule were identified from the literature and investigated (see Section **2.2.1**). The primary focus of investigation was to assess the importance of zero point motion in the system, this involved studying the system at the DFT level as presented here and the PIMD as discussed in Chapter **5**. Due to the relative cheapness of the DFT simulations several small investigations were carried out concurrently and they will also be reported here.

One such configuration was labelled III-i-1 by Re *et al.* [**60**] it is illustrated here in Figure **4.7a** and denoted config H. Our DFT simulations at 300 K show that the proton labelled H1 moves with considerable freedom between oxygens O1 and O5. Furthermore, Figure **4.7b** demonstrates an anticorrelation between the length R_c of the dissociating bond O1-H1 and the sum of the lengths of the neighbouring hydrogen bonds, labelled O3-H7 and O4-H6 in Figure **4.7a**, and denoted R_{hy}. The formation of the 'ionised' state due to the switch to the O5-H1 bond (such that the value of R_c is large) is seen to depend upon the prior existence of both the neighbouring hydrogen bonds (namely a low value of R_{hy}). If either neighbouring hydrogen bond is broken the system remains 'neutral' (with a low value of R_c), which is not surprising since the configuration is then similar to the SATH structure shown in Figure **4.5a**. This is an important corollary to conclusions acquired from consideration of geometry optimisation at 0 K, where config H has been shown to ionise [**60**].



(a)



(b)

Figure 4.7: The configuration denoted config H is shown in (A) with labels that identify certain O-H pairs. Plot (B) illustrates the probability density (given in arbitrary units) as a function of two structural features labelled R_c (the length of the covalent bond O1-H1) and R_{hy} (the sum of the lengths of prospective hydrogen bonds O4-H6 and O3-H7), obtained at DFT level. The associated potential of mean force takes the form of a broad, shallow well where the ionisation of the configuration is correlated with the status of the adjacent hydrogen bonds, as denoted by R_{hy}.

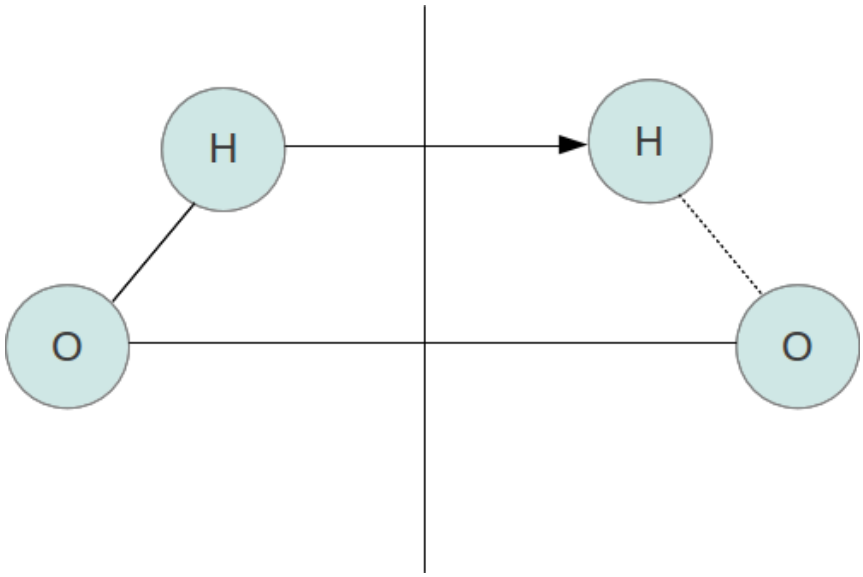
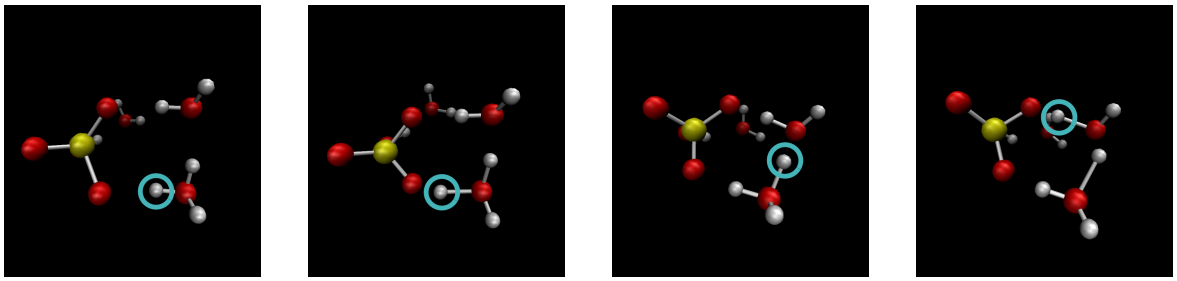


Figure 4.8: The transformation performed when the proton transfer event is forced. The proton was transferred from the sulphuric acid to a water molecule to which it was hydrogen bonded. The remaining atomic nuclei of the system are left in the state found by a geometry optimisation run.

At 300 K the behaviour can most certainly not be represented by harmonic fluctuations about an ionised mean structure and a free energy based on the rigid-rotor-harmonic-approximation for this configuration would be questionable due to significant anharmonic contributions.

Although not studied in any depth it is important to note that structural changes and proton transfer were observed numerous times in simulations which gives evidence of the liquid like behaviour of these clusters within the limited time periods studied. No evaporation event was observed, this is thought to be due to the length of these simulations. Another important question is the stability of the system once the proton transfer event has occurred. The neutral system was forced to undergo proton transfer in order to determine the existence of a barrier between the two steps. The method for forcing the proton transfer was to take the geometry optimised neutral structure, and then to reflect the hydrogen nucleus position (this proton is bonded to the sulphuric acid molecule and hydrogen bonded to a water molecule) in the plane defined to be perpendicular to the oxygen-oxygen vector of the hydrogen bond, and passes through the half way point on the same oxygen-oxygen vector. Figure **4.8** depicts the process of forcing proton transfer in the hydrogen bond system.

Both hb1 and hb2 as shown in Figure **4.5** were forced to perform transfer events in this way. Both systems became neutral in the order of less than 100 time-steps (100 fs). The problem with this method is that in this time the system is not in equilibrium. Unfortunately this means the system can act artificially and not give true dynamics. However the speed of the neutralisation suggests that the



(a) (b) (c) (d)

Figure 4.9: The sequence of events which occurred after forcing the proton to transfer in the hydrogen bond labelled as hb2.

ionised states are unstable which agrees with the geometry optimisations. Interestingly the method in which the system becomes neutral is different for hb1 and hb2, this is probably just pure luck but it gives a good opportunity to study a chain of proton transfer events within the sulphuric acid and water system.

The ionised hb1 system became neutral as the proton jumped back to the sulphuric acid through the hydrogen bond within a few time-steps. The system then remained stable for the rest of the simulation. hb2 was more interesting in that it did not follow the same route. Instead of that seen in hb1 the process that neutralised hb2 resembles the Grotthuss mechanism **[80]** where an excess proton is transferred across a water network whilst taking advantage of the ability to switch the proton which is being transferred between water molecules. Essentially this means that the system can switch which proton[7] is defined as the excess proton within the system given it more freedom to transfer a proton between molecules. Figure **4.9** is a series of screen shots outlining the major events in the forced proton transfer event. Figure **4.9a** is the initial set up with a hydronium. The proton travels between the sulphuric acid and the water several times. Eventually the proton becomes stable but at the cost of the hydrogen involved in the other hydrogen bond with the water becoming unstable. Figure **4.9c** shows that the proton is now attached to the second water in the two water loop. The system becomes neutral by returning the hydrogen in the second water to the sulphuric acid as shown in Figure **4.9d**. This configuration is essentially identical to the original neutral configuration. However, it has shuttled the proton around the loop rather than returning the proton at the original location. This process takes 0.1 ps to finish, and the system is then stable in the new configuration for the remaining 3.9 ps of simulation time.

[7]This freedom is for the structure the excess proton is attached to. This can be the hydronium ion, $[H_3O]^+$, zundel cation, $[H_5O_2]^+$ or an eigen cation, $[H_7O_3]^+$.

We now turn our attention to the question of the importance of the quantum nature of the nuclei in these clusters of sulphuric acid and water. The path integral molecular dynamics method is used in the next chapter along with DFT to address this question.

Chapter 5

Path integral molecular dynamics

This chapter introduces the path integral molecular dynamics method which allows the sampling of several possible paths taken by a particle. This effectively allows for the introduction of quantum behaviour within a classical framework. The method is then used to extend the work from the previous section by introducing zero-point motion to the nuclear cores of small clusters of sulphuric acid and water using the PIMD technique.

5.1 Introduction

The probability, $\mathcal{P}$, of a particle in the state[1] x at time t_1 being in the state x' at time t_2 is equivalent to the modulus squared of the sum of the amplitudes, $A(x)$ of all possible paths between the two states, i.e.

$$\mathcal{P} = \left| \sum_{paths} A_{path}(x', t_2; x, t_1) \right|^2 \tag{5.1.1}$$

The paths must be considered in this way as there are interference terms which cannot be neglected when considering the system quantum mechanically. A well understood example is the double slit experiment [**12**, **88**] where the observed intensity of electrons passing through the slits is strong evidence for each electron passing through both slits. The detected intensity after the slits is well described by considering the interference of the two paths (each path is a route through a slit to the detector). The amplitude for a path from the state x to the state x' in the time interval $t = t_2 - t_1$ is

[1] The x expression for the state is deliberate as the state is an eigenfunction of the position operator.

given by

$$A = \left\langle x' \left| e^{-i\hat{H}t/\hbar} \right| x \right\rangle \tag{5.1.2}$$

The expression is equivalent to the time evolution operator $\exp(-i\hat{H}t/\hbar)$ acting on state x projected on the state x' integrated over the volume. It is convenient to note the relationship of the time evolution operator and the canonical density matrix, $\hat{\rho}$, given by

$$\hat{\rho}(\beta) = e^{-\beta\hat{H}} \tag{5.1.3}$$

it is clear that $\hat{\rho}(it/\hbar) = \exp(-i\hat{H}t/\hbar)$, in other words it is possible to calculate the time evolution operator by evaluating the canonical density matrix at an imaginary time $t = -i\beta\hbar$. This transformation is known as a Wick rotation **[12]**. In general this involves switching between real and imaginary parameters in order to reduce the difficulty in obtaining a solution to a given problem. Indeed in this case the Wick rotation is used to derive the Feynman path integral for the canonical density matrix instead of the time evolution operator.

We continue the derivation by defining the kinetic energy operator, $\hat{T}$, potential energy operator, $\hat{U}$ and the Hamiltonian which is the sum of the two previous operators, $\hat{H} = \hat{T} + \hat{U}$. These operators do not in general commute, i.e. $[\hat{T}, \hat{U}] \neq 0$ which complicates the evaluation of properties based on this quantity. Fortunately the Trotter theorem can be used to simplify the maths involved with using the $\hat{H}$ expression. The Trotter theorem **[139]** states that for two non commuting operators $\hat{L}_1$ and $\hat{L}_2$ we find,

$$e^{\hat{L}_1+\hat{L}_2} = \lim_{P\to\infty} \left[e^{\hat{L}_1/2P} e^{\hat{L}_2/P} e^{\hat{L}_1/2P} \right]^P \tag{5.1.4}$$

Defining the quantity $\hat{\mathcal{H}} \equiv e^{-\beta\hat{U}/2P} e^{-\beta\hat{T}/P} e^{-\beta\hat{U}/2P}$ we find the following expression for elements of $\hat{\rho}(it/\hbar)$ which are related to Eq. (**5.1.2**) via a Wick transform to the canonical density matrix and an expansion using the Trotter theorem,

$$\rho(x, x'; \beta) = \lim_{P\to\infty} \left\langle x' \left| \hat{\mathcal{H}}^P \right| x \right\rangle \tag{5.1.5}$$

The expression is simplified by inserting the identity operator, $\hat{I} = \int dx \, |x\rangle \langle x|$ between each $\hat{\mathcal{H}}$ in Eq. (**5.1.5**). This will require $P - 1$ identity operators and so to avoid confusion with x_1 and x these will

be labelled from x_2 onwards such that,

$$\rho(x, x'; \beta) = \lim_{P\to\infty} \int dx_{2\ldots} dx_P \left\langle x' \left| \hat{\mathcal{H}} \right| x_P \right\rangle \left\langle x_P \left| \hat{\mathcal{H}} \right| x_{P-1} \right\rangle \ldots \left\langle x_2 \left| \hat{\mathcal{H}} \right| x \right\rangle \tag{5.1.6}$$

Conveniently the potential energy operator $\hat{U}$ is an eigenfunction of the state $|x\rangle$, so that Eq. (**5.1.6**) can be simplified by the following observation,

$$\left\langle x_{i+1} \left| \mathcal{H} \right| x_i \right\rangle = e^{-\beta U(x_{i+1})/2P} \left\langle x_{i+1} \left| e^{-\beta \hat{T}/P} \right| x_i \right\rangle e^{-\beta U(x_i)/2P} \tag{5.1.7}$$

Unfortunately the same cannot be done for the kinetic operator. However a similar trick can be used in which a different identity operator $\hat{I} = \int dp \, |p\rangle \langle p|$ is used. Insert this into Eq. (**5.1.7**) between the kinetic energy operator and state $|x_i\rangle$ it is then possible to derive the following expression,

$$\rho(x, x'; \beta) = \lim_{P\to\infty} \left(\frac{mP}{2\pi\beta\hbar^2} \right) \int dx_2 \ldots dx_P \exp \left[-\frac{1}{\hbar} \sum_{i=1}^{P} \left\{ \frac{mP}{2\beta\hbar} (x_{i+1} - x_i)^2 + \frac{\beta\hbar}{2P} (U(x_i) + U(x_{i+1})) \right\} \right] \Bigg|_{x_1=x}^{x_{P+1}=x'} \tag{5.1.8}$$

here m refers to the mass of the particle. It is useful to introduce the notion of beads which define a possible position of a particle in state $|x_i\rangle$. Each bead has interaction with other particles in state $|x_i\rangle$ and a spring interaction with the neighbouring state's representation of that bead as seen from Eq. (**5.1.8**).

It is possible to obtain the canonical partition function $Q(L, T)$ from Eq. (**5.1.8**) where L defines the state space, i.e. $x \in [0, L]$ and T is the temperature. The canonical partition function is defined as the trace of the canonical density matrix, $Q(L, T) = tr[\rho]$, and can be found by setting $x_{P+1} = x_1$ so that the sum is considered closed. The resulting expression is simplified by observing that $\sum_{i=1}^{N} (x_i + x_{i+1})/2 = \sum_{i=1}^{N} x_i$ under the condition that $x_{N+1} = x_1$. In addition the first term in the exponential is similar to the form of a harmonic spring potential. Using this analogy it is possible to define a spring constant which quantifies the vibration between adjacent degrees of freedom with an angular frequency, ω_P, defined by $\omega_P = \sqrt{P}/\beta\hbar$. It is clear that the stiffness of this spring is dependent on the number of beads in the simulation. The canonical partition function is therefore of

the form,

$$Q(L,T) = \lim_{P\to\infty}\left(\frac{mP}{2\pi\beta\hbar^2}\right)\int dx_1\ldots dx_P \exp\left[-\frac{1}{\hbar}\sum_{i=1}^{P}\left\{\frac{m}{2}\beta\omega_P^2(x_{i+1}-x_i)^2+\frac{\beta\hbar}{P}U(x_i)\right\}\right]\Bigg|_{x_{P+1}=x_1} \tag{5.1.9}$$

The equation has added an integral over the variable x_1. This represents the integral over the possible states in the system. The condition $x_{P+1} = x_1$ indicates that for each state x_1 the particle is allowed to explore paths of P steps before returning to its original state. The application of this technique in MD simulations is not yet clear. The next section develops a practical MD method from the canonical partition function derived here.

5.2 Molecular dynamics with path integrals

It is clear from Eq. (**5.1.9**) that canonical partition function has no momentum dependence which is a requirement for performing MD simulations. Therefore, the first step of forming a molecular dynamics technique is to add this dependence to Eq. (**5.1.9**). This will consist of adding a parameters which act as a set of conjugate momenta variables, p, to the x variables. This is justified by manipulating Eq. (**5.1.9**) to resemble the classical canonical partition function of a cyclic polymer chain in a classical potential $U(x_i)/P$ and recasting the prefactor as a set of Gaussian integrals over $p_1,\ldots,p_P$ [**12**]. Rearranging Eq. (**5.1.9**) with the addition of these momenta gives,

$$\begin{aligned} Q(L,T) &= \lim_{P\to\infty}\left(\frac{mP}{2\pi\beta\hbar^2}\right)\int dp_1\ldots dp_P \\ &\quad \int dx_1\ldots dx_P \exp\left[-\beta\sum_{i=1}^{P}\left\{\frac{p_i^2}{2m'}+\frac{m}{2}\omega_P^2(x_{i+1}-x_i)^2+\frac{\beta\hbar}{P}U(x_i)\right\}\right]\Bigg|_{x_{P+1}=x_1} \end{aligned} \tag{5.2.1}$$

Introducing the momentum parameter p_i also requires a mass term m' which is formally given by $m' = mP/(2\pi\hbar)^2$. This new mass parameter does not affect equilibrium averages and as such is effectively a free parameter in the system. This has important consequences in Section **5.3**. The next step is to use a "classical isomorphism" to define the classical Hamiltonian with a finite P such that,

$$H_{cl}(x,p) = \sum_{i=1}^{P}\left\{\frac{p_i^2}{2m'}+\frac{m}{2}\omega_P^2(x_{i+1}-x_i)^2+\frac{\beta\hbar}{P}U(x_i)\right\} \tag{5.2.2}$$

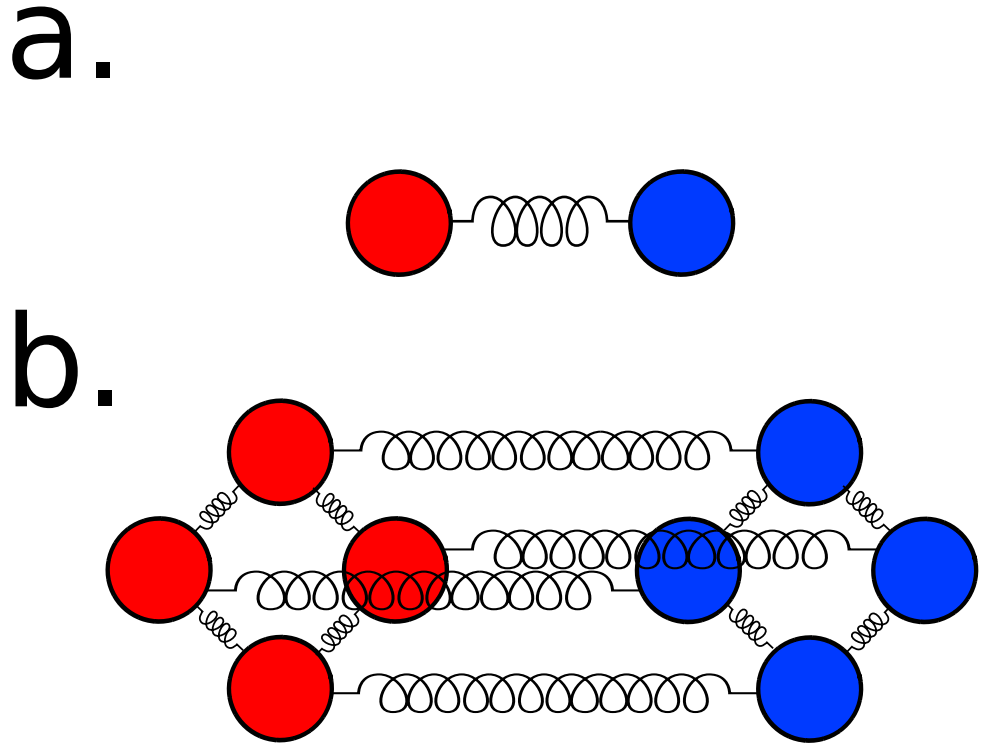


Figure 5.1: A diagram showing the PIMD representing of a two interacting particles labelled red and blue. a. and b. refers to the classical and the PIMD simulation with four beads respectively. Figure **b** shows the additional interactions to neighbouring beads as introduced in the PIMD method.

A molecular dynamics scheme can be derived from the Hamiltonian which yields the following equations of motion,

$$\dot{x}_i = \frac{p_i}{m'} \tag{5.2.3}$$

$$\dot{p}_i = -m\omega_P(2x_i - x_{i+1} - x_{i-1}) - \frac{1}{P}\frac{\partial U(x_i)}{\partial x_i}$$

This resembles a set of beads, connected by harmonic springs in the presence of an external potential. Figure **5.1** shows a four bead representation of two interacting particles. Each bead shows a possible path that the particle it represents can take.

5.3 Staging transformation

When implementing the PIMD technique there are issues which affect the convergence [**140**]. These are due to the range of time scales introduced by the inter-bead harmonic springs (ω_P). Figure **5.2** shows two different modes in a three bead simulation. The range of vibrations introduced are linked with the inter-bead potential making it difficult to sample the entire range efficiently. A transformation of variables can improve the treatment of this time scale problem. There are two

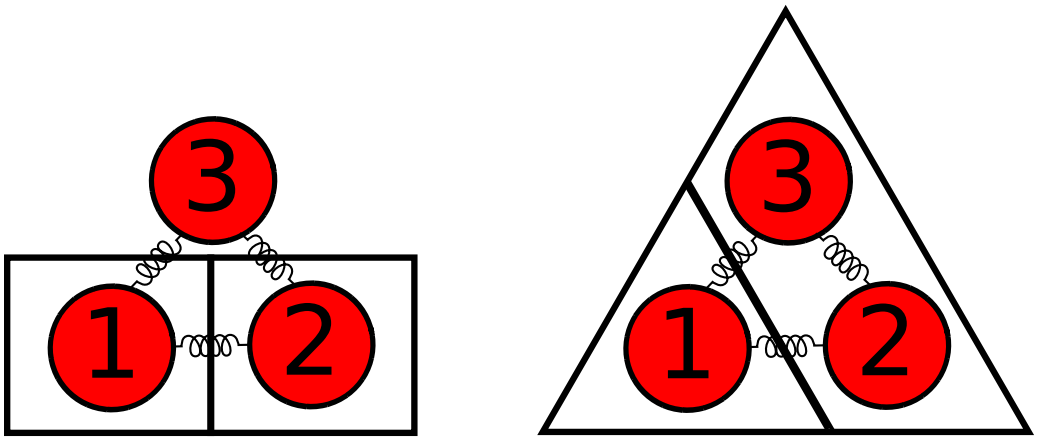


Figure 5.2: A three bead representation of a particle with interconnecting springs. Each spring vibrates with a different frequency ω_P, additionally there are different modes of vibrations. Two examples are depicted here, the first between bead one and two, and the second between bead one and beads two and three. Efficient sampling of these vibrations is difficult due to the introduced range of frequencies.

common transformations known as normal mode [**141**, **142**] and staging transformation [**142**, **143**]. In this project the staging transformation is used and a brief description of the method is presented here. The staging transformation makes the following inverse transformation to the x_i variable,

$$\begin{aligned} x_1 &= u_1 \\ x_i &= u_1 + \textstyle\sum_{l=i}^{P} \frac{i-1}{l-1} u_l \end{aligned} \tag{5.3.1}$$

In addition the mass terms m' and m are replaced with a set of masses. The main effect of this transformation on Eq. (**5.2.1**) is that the harmonic coupling has been separated such that,

$$Q_P(L,T) = \int dp_1 \dots dp_P \int du_1 \dots du_P \exp\left\{ -\beta \sum_{i=1}^{P} \left[\frac{p_i^2}{2m_i'} + \frac{1}{2} m_i \omega_P^2 u_i^2 + \frac{1}{P} U(x_i(u_i)) \right] \right\} \tag{5.3.2}$$

where $m_1 = 0$ and $m_i = \frac{i}{i-1} m$ for $i = 2, \dots, P$. The term m_i' is given by m for $i = 1$, otherwise it is given by m_i. $x_i(u)$ is the inverse transformation for x_i as given in Eq. (**5.3.1**). It is clear that the first term in the sum has no harmonic spring term. This term represents the entire chain of beads i.e. the particle. The spring term has been completely uncoupled meaning that independent frequency sampling can be made. The equation of motion derived from Eq. (**5.3.2**) is given by,

$$\begin{aligned} \dot{u}_i &= \tfrac{p_i}{m_i'} \\ \dot{p}_i &= -m_i \omega_P^2 u_i - \tfrac{1}{P} \tfrac{\partial U}{\partial u_i} \end{aligned} \tag{5.3.3}$$

Further, it is useful to expand the partial derivative term using the chain rule in the following way,

$$\begin{aligned}\frac{1}{P}\frac{\partial U}{\partial u_1} &= \frac{1}{P}\sum_{l=1}^{P}\frac{\partial U}{\partial x_l} \\ \frac{1}{P}\frac{\partial U}{\partial u_i} &= \frac{1}{P}\left[\frac{\partial U}{\partial x_i} + \frac{i-2}{i-1}\frac{\partial U}{\partial u_{i-1}}\right]\end{aligned} \tag{5.3.4}$$

This has the benefit of expressing the partial derivative of the external potential in the original coordinate system which can generally be computed given the form of the external potential U. In addition each degree of motion (for each bead) needs to be attached to a thermostat to ensure a canonical distribution.

The PIMD allows the expansion of Eq. (**4.1.8**) by sampling P-step paths in the quantum canonical partition function. A staging transformation allows the decoupling of the harmonic spring terms which have been introduced. A thermostat is coupled to each degree of freedom for each bead which ensures the canonical distribution. In addition the form is now suitable for a parallel environment as each time-step for bead can be calculated on a separate set of cpu's essentially meaning that in a parallel environment the additional time cost can be interchanged for additional cpu's. The Langevin thermostat is used in this work and is described in the Section **3.5**.

5.4 Simulation details

The path integral molecular dynamics simulations follows immediately from the electronic DFT simulations discussed in the previous chapter and details on this underlying electronic potential are described in Section **4.2**. Continuing with this work, we studied configurations of $[H_2SO_4][H_2O]_{n=3-4}$ at a target temperature of 300 K with the CASTEP [**127**] (version 5.5) code. The CASTEP code uses the PIMD method to approximate the zero-point motion contributions to the nuclear cores. Simulation were ran with P beads where $P = 1$, 4, 8, 16 and 32. The staging transformation [**12**] was used for all PIMD simulations. The $P = 1$ case represents the classical limit of the PIMD technique and corresponds to the complete neglect of zero-point motion. For details of the simulation lengths and time-step refer to Table **5.1**.

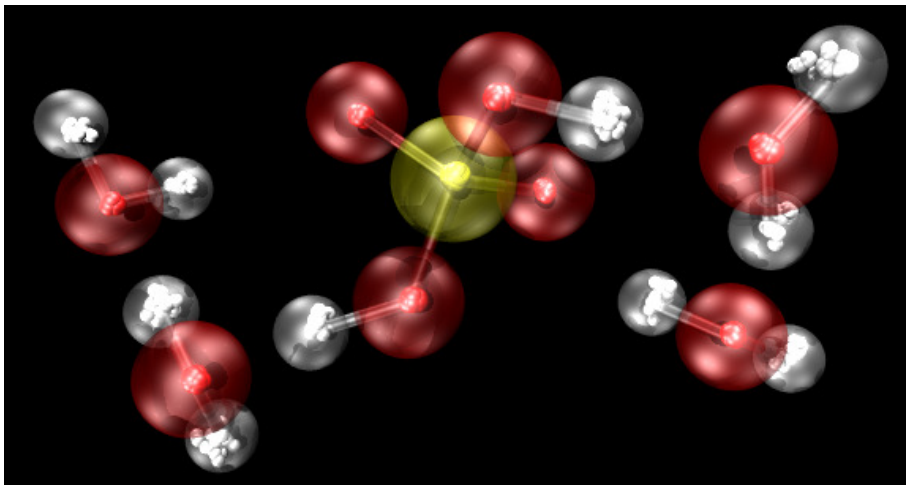


Figure 5.3: A 16 bead representation (each atom in the system is represent by 16 beads) of a system containing one sulphuric acid and four water molecules: the distribution of bead positions conveys the quantum uncertainty. One set of beads has been represented twice, first as a small beads along with the 15 other sets of beads, and secondly as a large translucent sphere which gives an outline of the configuration of the system.

Figure **5.3** is a snapshot from a 16 bead simulation representing the behaviour of a cluster of one sulphuric acid and four water molecules. The spatial separation of the beads clearly illustrates the greater positional uncertainty of the hydrogen nuclei compared to that of the oxygen and the sulphur nuclei.

5.5 Results

A PIMD study was performed first for two low energy configurations (denoted SATH and SAQH) identified in the literature **[58, 60]** and shown in Figures **4.5a** and **4.5b**. It is envisaged that hydrogen bonds, in particular those associated with the sulphuric acid, would be the most susceptible to zero-point effects due to the inherent tendency of sulphuric acid to dissociate. Figure **5.4** shows the average oxygen-oxygen distance (d_{OO}) of specific hydrogen bonds as a function of the number of beads representing atoms in the system. The bonds labelled hb1 and hb2 in the SATH structure contract in length by around $2-5\%$ with respect to the outcome of classical dynamics while the situation for hb3 is less clear. Note that the longest simulations were performed for the single bead and 16 bead representations of the SATH structure, as indicated in Table **5.1**. For other cases shorter studies were performed to illustrate the trends, though the accuracy of the results is lower. The overall picture from these result indicate that zero point motion has a minor contribution in these

	time-step [fs]	Simulation time [ps]
SATH 1 bead	1.00	11.000
SATH 4 bead	0.50	1.500
SATH 8 bead [2]	0.25	0.875
SATH 16 bead	0.50	10.673
SATH 32 bead	0.50	0.512
SAQH 1 bead	1.00	1.000
SAQH 4 bead	0.50	1.500
SAQH 8 bead	0.50	1.500
SAQH 16 bead	0.50	1.500
config H 1 bead	1.00	10.900
config H 16 bead	1.00	10.647

Table 5.1: Compilation of the simulation length and time-step for the MD runs performed. SATH refers to sulphuric acid trihydrate and SAQH refers to sulphuric acid tetrahydrate and correspond to structures shown in Figures **4.5a** and **4.5b** respectively. Config H refers to the trihydrate configuration shown in Figure **4.7a**. Note that the longest simulations were performed for SATH and config H. The 1 bead simulations are the same as those described in Table **4.1**.

clusters at 300 K.

Next we examine in detail how the behaviour of the hydrogen atom in hydrogen bond hb2 is affected by PIMD. This is explored by constructing a potential of mean force (PMF, see Section **3.8.2**) for the hydrogen, defined by:

$$W(R,\beta) = -k_B T \ln g(R,\beta)$$

where R and β are geometric parameters illustrated in Figure **3.6** and $g(R,\beta)$ is the proportion of simulation snapshots with the hydrogen located within the region defined by $R \rightarrow R + dR$ and $\beta \rightarrow \beta + d\beta$ divided by the equivalent proportion for noninteracting particles. For the PIMD simulations the centroid of the beads representing the hydrogen atom was used to produce the PMF. The method is described extensively in reference [**122**]. Figures **5.5** and **5.7** show the PMFs acquired using classical MD and PIMD, respectively, for hydrogen bonds hb1 and hb2. For the hb1 statistics a condition was applied such that the water molecule attached to the sulphuric acid must not be hydrogen bonded to another water, this was achieved by applying a distance criteria to the data. For hb2 a similar method was used in reverse to ensure that the water was hydrogen bonded to a further

[2]Initially it was thought necessary to halve the time step for every doubling in the number of beads due to the relationship between the number of beads and the strength of the interbead spring (see Eq. (**5.3.2**)). However, the computation demand of the simulations meant that time steps of was used for future simulations involving four or more beads. Fortunately this was found to give stable simulations in comparison to simulations ran with a longer time-step of 2 fs which was found to make the configuration unstable and give rise to an erratic energy landscape.

water molecule. This rudimentary condition was used to ensure that the bonding configuration that was expected was present in the particular time step of the simulation as the hydrogen bonds were prone to breaking.

For hb1 there is very little noticeable difference between the classical simulation and that of the PIMD simulation with 16 beads. Figure **5.6** shows a single dimensional representation of Figures **5.5a** and **5.5b** achieved by integrating over the β parameter. This shows a strong agreement between the two simulations leading to the conclusion that for this bond the contribution of zero-point motion is small.

The PMF plots in Figure **5.7** visualise the differences between the dynamics of the hb2 bond in Figure **4.5a** under classical MD and the PIMD schemes. Such a comparison is limited by the computationally expensive techniques employed. However it does offer an insight into the importance of zero-point effects in small clusters of sulphuric acid and water. The main effect is a shift in the minimum of the PMF of hydrogen bond length R by about 0.2 Å going from the DFT to the PIMD result indicating that the zero-point motion has a mean configurational influence on this bond. Figure **5.8** is a one dimensional version of Figures **5.7a** and **5.7a** obtained by integrating over the β parameter, this makes the shift in the R parameter more apparent and implies that the PIMD case has shortened the hydrogen bond over the standard DFT case. This suggests that the introduced zero-point motion has increased the likelihood of proton transfer in this particular bond in this system. Ionisation of the cluster is expected to increase the nucleation rate of the sulphuric acid and water system (see Section **2.1**) and so an increased propensity for ionisation could have implications for the nucleation rate calculated from quantum chemistry methods such as the popular DFT geometry optimisation (see Section **2.2.1**).

The effects of zero-point motion are clearly rather subtle. To explore this further, we return to the delicate switching behaviour of the O1-H1-O5 bonds discussed in Section **4.3** and contrast the classical and PIMD results. Figure **5.9** illustrates the motion of the proton between the neutral and ionised positions, discussed earlier, in terms of the O1-H1 bond length. Which of the nuclei O1 or O5 was the nearest neighbour to the H1 nucleus (see Figure **4.7a**) was monitored to quantify this hopping behaviour. It was found that in the classical case H1 was closer to O1 for 21.5% percent of the simulation with standard error[3] $\sigma_{SE} = 3.2\%$ whereas in the 16 bead PIMD simulation this

[3]the standard errors were obtained using the standard blocking procedure [**108**, **121**].

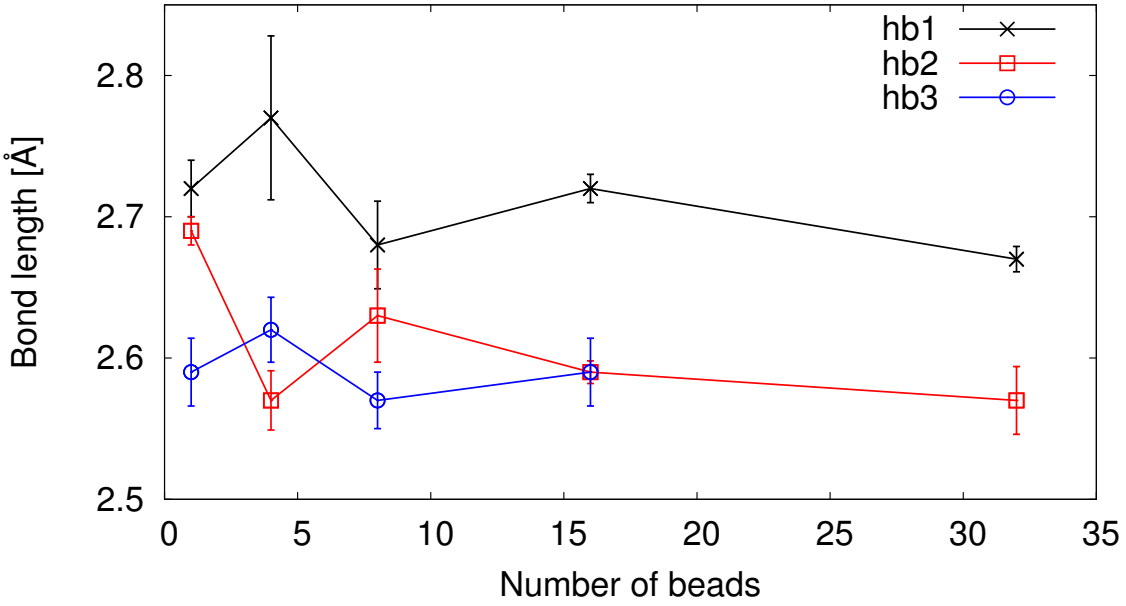
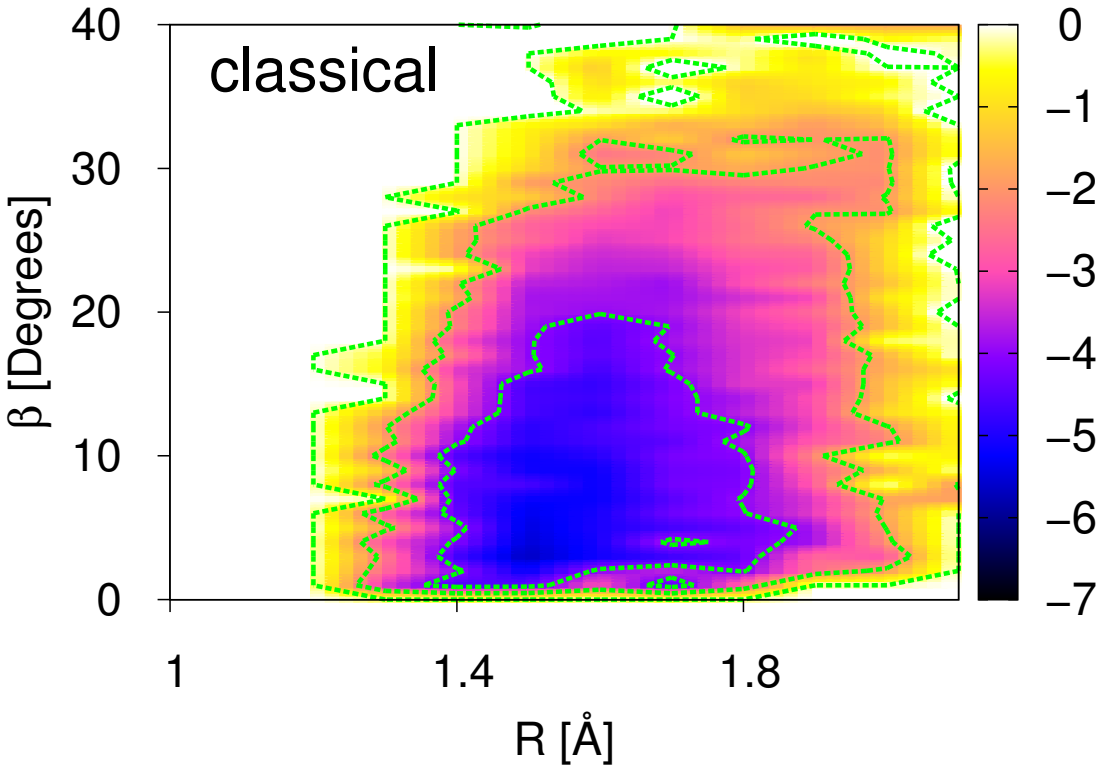
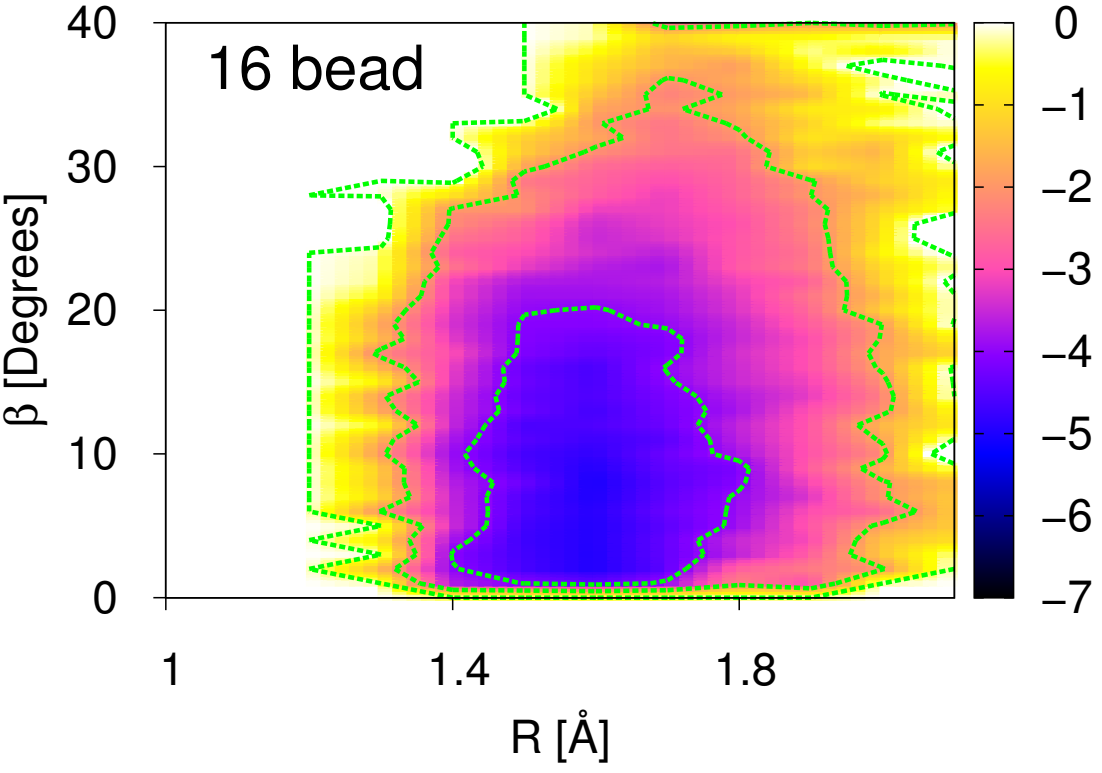


Figure 5.4: The average oxygen-oxygen separation d_{OO} of specific hydrogen bonds as a function of the number of beads used in the simulation. Labels hb1 and hb2 refer to Figure **4.5a** and hb3 is shown in Figure **4.5b**. The error bars were determined by the standard blocking procedure [**108**, **121**] and a blocking length of 0.256 ps was found to give independent sampling. The calculations correspond to the cases listed in Table **5.1**. A limit on the computational power available resulted in no 32 bead simulation for hb3.

figure dropped to 14.8% with $\sigma_{SE} = 2.7\%$. This property was further investigated by defining a threshold for the O1-H1 bond length below which the system is considered neutral, and beyond which it is better described as ionised. We define a 1.22 Å distance to separate the two regimes, and this is shown as a horizontal line in Figure **5.9**. For the classical dynamics, the percentage of time the system remains neutral according to this criterion is 20.1% with $\sigma_{SE} = 2.9\%$. An analysis of the PIMD simulation with 16 beads yields a corresponding percentage of neutral residence time of 12.5% with $\sigma_{SE} = 2.4\%$. These results are consistent with those determined from the nearest neighbour criterion. The proportion of time spent in the ionised configuration rises from 79.1% to 87.5%. This suggests that the inclusion of zero-point motion promotes the formation of the ionised state; quantum uncertainty favours proton transfer. The standard errors presented here is of concern as they are of order of the differences in which the state is described as ionised and that when the state is described as neutral (according to the relevant criterion). This can be addressed by additional simulations of this system, however time and computation demands meant this was not possible in the study performed here.



(a)



(b)

Figure 5.5: Contour plots of the potential of mean force $W(R, \beta)$ in units of $k_B T$ for the hydrogen in the bond labelled hb1 in Figure **4.5**. The green dashed lines indicate contour levels of 0, -2, -4 and -6. Plot (A) shows results from standard DFT molecular dynamics and plot (B) arises from PIMD using 16 beads. The simulation times are given in Table **5.1**.

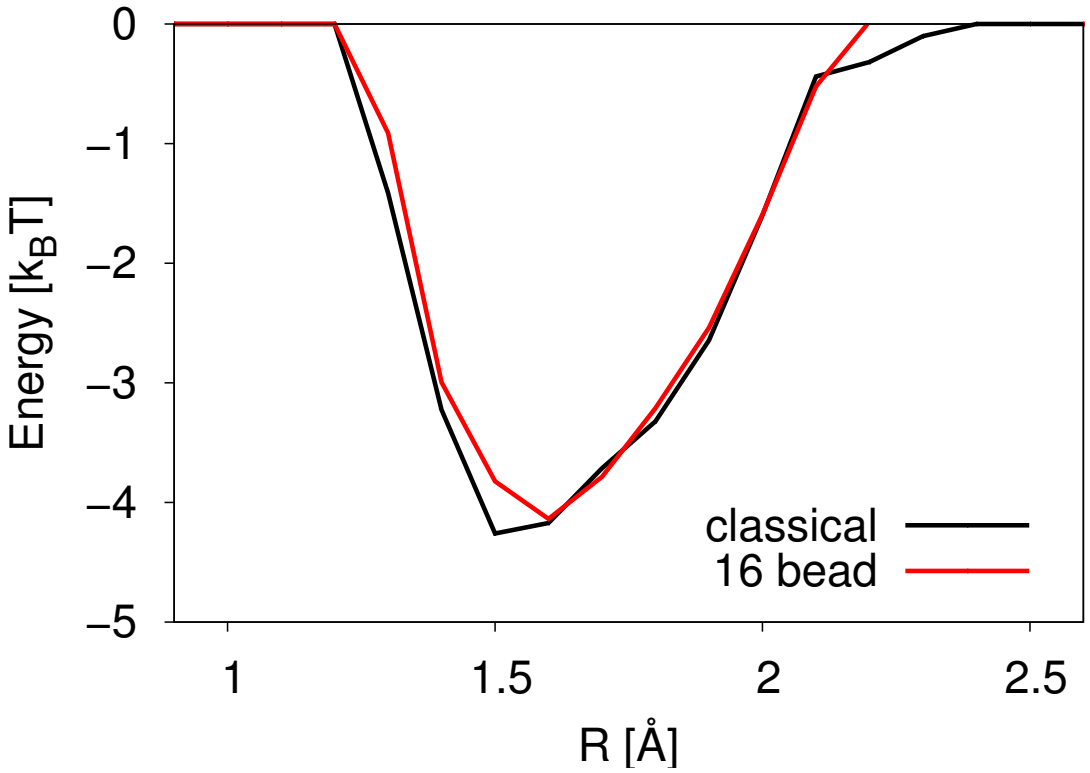
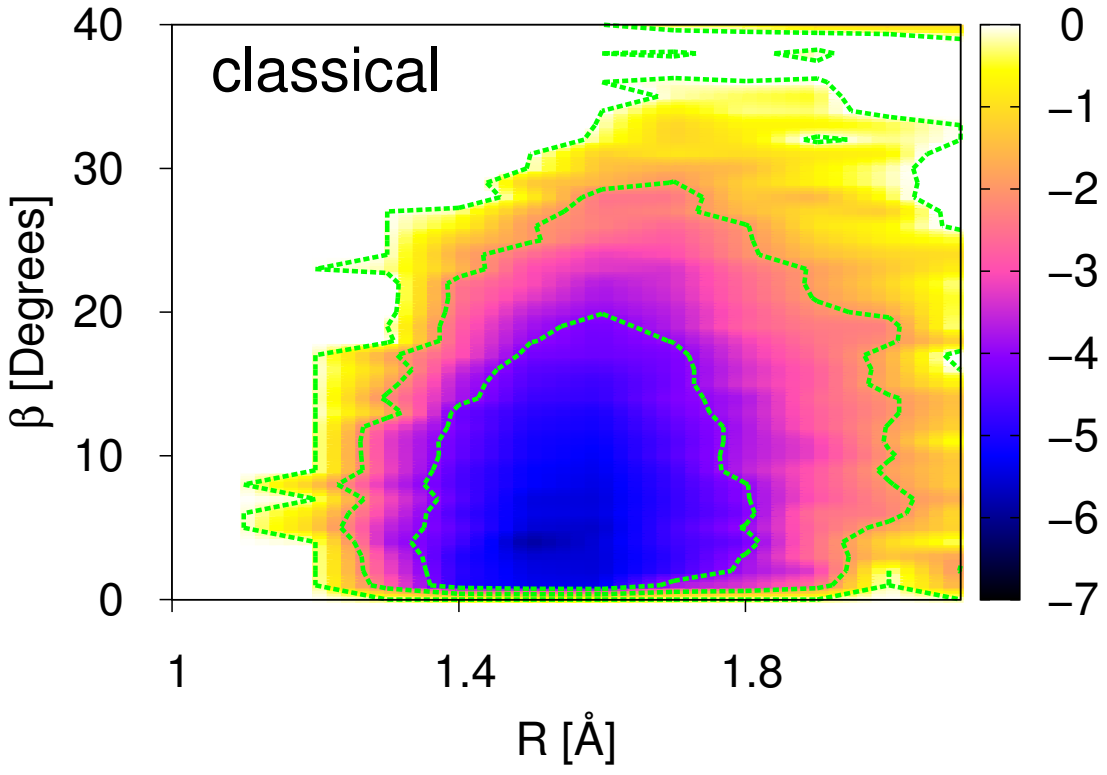


Figure 5.6: Plot shows a 1D version of Figure **5.5** obtained by integrating over the β parameter.

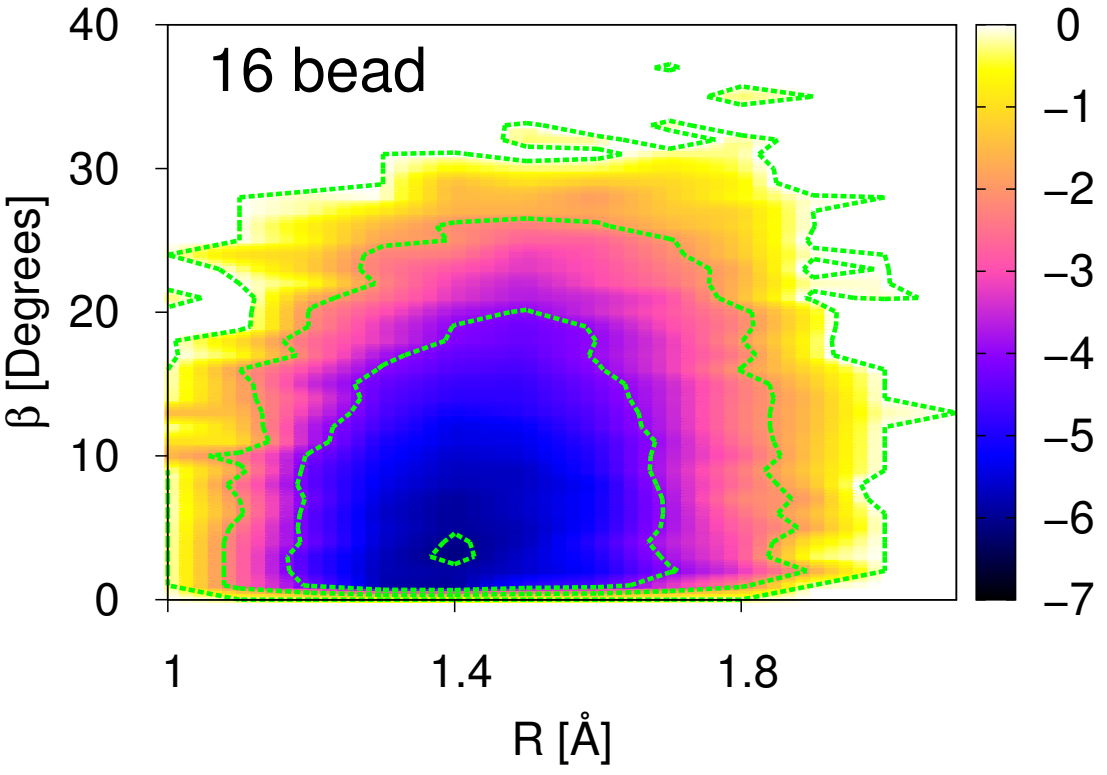
5.6 Conclusion

This study of small clusters of water and sulphuric acid molecules leads us to two main conclusions. Firstly, it has been demonstrated that molecular dynamics can reveal features that are not available from knowledge of the geometry optimised structure at zero temperature. The prime example of this is the complex behaviour of cluster configuration III-i-1 identified by Re *et al.* **[60]** and here denoted config H. This structure has been regarded as the most stable ionised configuration for the trihydrated sulphuric acid molecule **[58, 60, 69]**, but the results gained in this chapter and in Chapter **4** indicate that the structure exhibits both neutral and ionised characteristics at 300 K. This conclusion highlights the limitations of the rigid rotor harmonic approximation **[144, 145]** for free energy estimation using a single optimised structure.

Secondly, the inclusion of zero-point motion through PIMD simulations of the sulphuric acid-water system has been shown to produce small but clear structural distortion at 300 K in a selected number of configurations when compared with classical dynamics. The mean oxygen-oxygen separation of hydrogen bonds hb1 and hb2 in the structure shown in Figure **4.5a** is reduced by $2-5\%$. A mild shortening of the hb2 hydrogen bond length was observed, shown by constructing



(a)



(b)

Figure 5.7: Contour plots of the potential of mean force $W(R, \beta)$ in units of $k_B T$ for the hydrogen in the bond labelled hb2 in Figure **4.5**. The green dashed lines indicate contour levels of 0, -2, -4 and -6. Plot (A) shows results from standard DFT molecular dynamics and plot (B) arises from PIMD using 16 beads. The simulation times are given in Table **5.1**.

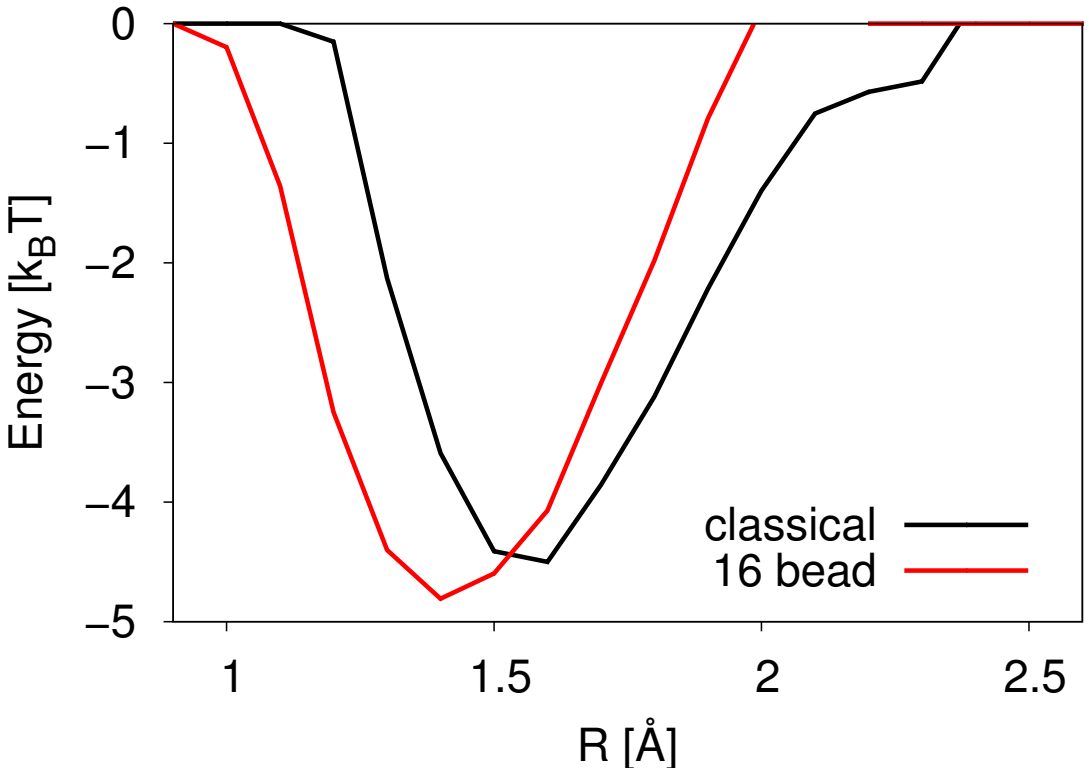


Figure 5.8: Plot shows a 1D version of Figure **5.7** obtained by integrating over the β parameter.

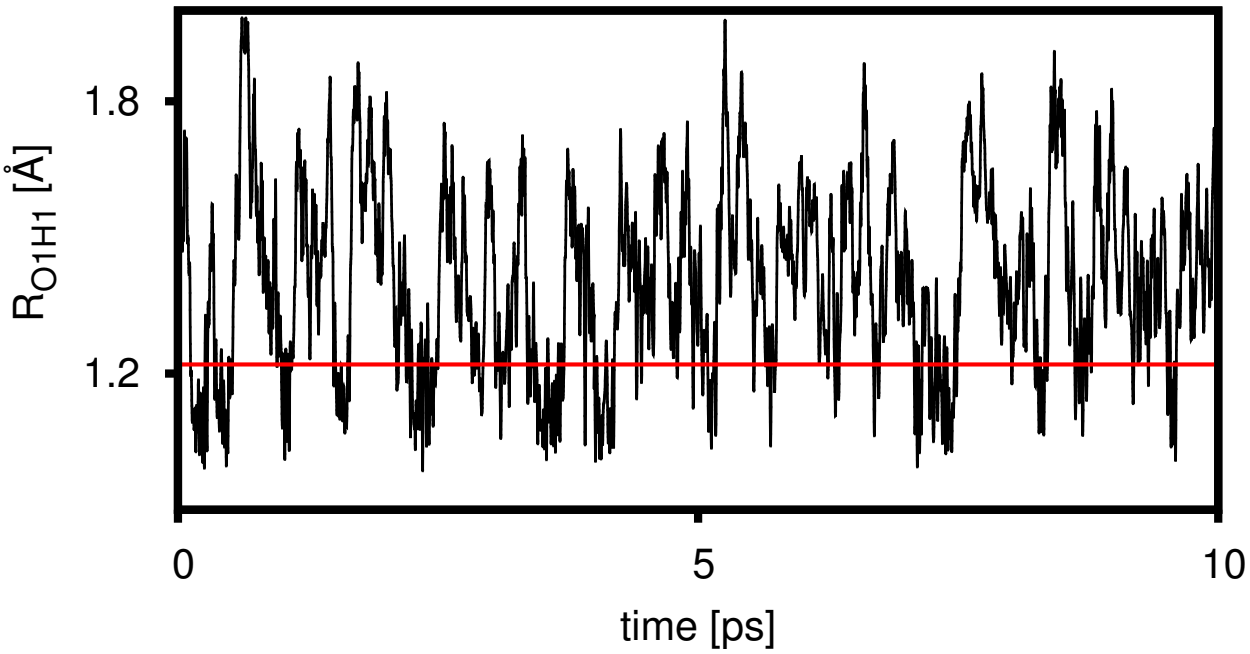


Figure 5.9: Plot of the O1-H1 bond length (denoted as R_{O1H1}) against time in config H in Figure **4.7a** from the 1 bead simulation as detailed in Table **5.1**. The plot clearly illustrates the motion between the neutral and ionised states. The horizontal line drawn at 1.22 Å provides a simple threshold between covalent and hydrogen bond-like behaviour of the O1-H1 bond.

potentials of mean force for the classical and PIMD schemes, as illustrated in Figure **5.7**, no significant affect was seen for hb1 as shown in Figure **5.6**. Furthermore, the results indicate that zero-point motion brings about a greater propensity for proton transfer in the O1-H1-O5 substructure of the configuration shown in Figure **4.7a** at 300 K.

This conclusion is consistent with the paper by Li *et al.* [**146**] where quantum nuclear effects[4] on the hydrogen bond are studied, specifically Figure 3 in reference [**146**] where the OO length is compared with the length of the projection of the covalent OH bond on the OO vector. The implication is that the projected covalent bond length is increased by quantum effects when the hydrogen bond is considered to be strong, as judged by a shift in vibrational frequency of the covalent OH bond due to the presence of the hydrogen bond.

The research supports the view that the zero-point effect is most significant in configurations where proton transfer is intrinsically likely. Classical DFT and PIMD simulations of the cluster shown in Figure **4.7a** have demonstrated frequent proton transfer. Using an O-H separation of 1.22 Å as a threshold for distinguishing the ionised from the neutral state, the cluster is found to remain neutral 20.1% of the time ($\sigma_{SE} = 2.9\%$) with classical MD and 12.5% ($\sigma_{SE} = 2.4\%$) according to PIMD. It is possible to infer that quantum effects have increased the degree of proton transfer. Simulations ran at a lower temperatures are expected to increase the significance of the zero-point effects, making this an avenue for future research. In addition, since substances such as ammonia and amines are increasingly thought to be relevant to atmospheric nucleation [**10**, **45**], assessing the importance of zero-point motion in these systems would also be of interest.

These simulations were run on 48 cpu cores for 10 hours in return for 1000 time-steps (the equivalent of 1 ps of simulation ran with a 1 fs time-step) of a one bead calculation. The project now continues with focus on the less expensive empirical approach to expand the simulation lengths available. The most difficult feature to retain from these higher level methods was the reactivity of the sulphuric acid and water system. As discussed in Section **2.4**, the empirical valence bond methodology was chosen to provide the framework for further studies. The development of the empirical potential scheme is described in the next chapter.

[4]The terminology of zero-point motion and quantum nuclear effects are interchangeable when discussing the quantum aspects of the behaviour of nuclei in this study.

Chapter 6

Empirical Valence Bond model

The choice of the EVB methodology [**87**] as a basis for designing a reactive empirical potential for future simulations is discussed in Section **2.4.1**. The EVB methodology was introduced in the 1980s by Warshel and Weiss [**99**] and progress was made by been made by Schmitt and Voth [**101**] who presented the MS-EVB model which is used to give a framework for the work presented din this thesis. This chapter gives a details of how the EVB methodology was applied to the sulphuric acid and water system.

6.1 Introduction

The EVB model constructs a potential energy surface of the system from a superposition of diabatic states[1], these are then referred to as basis states for simplicity. Each basis state describes a possible bonding configuration. In theory every possible[2] basis of the system should be considered, however in practice a large number of basis states make a negligible contribution to the potential energy of the system and thus can be ignored. An examples of which are shown in Figure **6.1** where there are two basis states of a system corresponding to the possibility that one of the hydrogen atoms can form part of either a sulphuric acid or a hydronium molecule. These have been chosen as they are the two largest contributions to the potential energy of the system. Once the basis states of the system are

[1] A diabatic state of the system in this case is refers to a state of the system that is not necessarily an eigenstate of the current Hamiltonian of the system, however, a superposition of basis states provides an eigenstate of the current Hamiltonian of the system. In practice the basis states generally refer to the reactant and product states of the system, the system is then described as superposition of the product and reactant states.

[2] In practice the EVB method is generally applied once the molecules of the system have been identified, for instance in the model presented here there are four molecules (sulphuric acid, bisulphate, water, and hydronium), this reduces the amount of basis states that are considered.

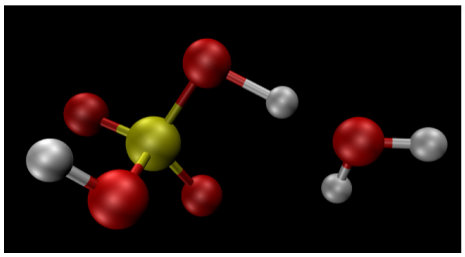
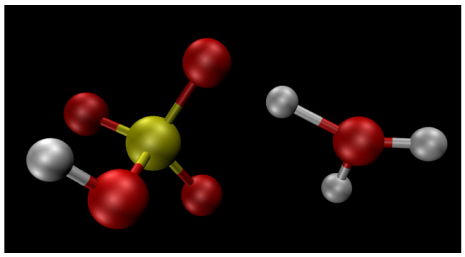

$|1\rangle$

$|2\rangle$

Figure 6.1: Images showing two possible basis states of bonding for the same configuration. State $|1\rangle$ contains a sulphuric acid and a water molecule whereas state $|2\rangle$ contains a bisulphate and a hydronium ion.

specified, a matrix $\mathcal{H}$ that resembles a Hamiltonian is constructed as follows,

$$\mathcal{H}_{ij} = \langle i| H |j\rangle \tag{6.1.1}$$

where $\langle i| H |i\rangle$ is the potential energy of the system according to the bonding configuration defined by $|i\rangle$ and an associated classical potential. $\langle i| H |j\rangle$ where $i \neq j$ represents the coupling between basis states $|i\rangle$ and $|j\rangle$. In practice $\langle i| H |j\rangle$ is described by an empirical function.

Once the $\mathcal{H}$ matrix has been constructed, the eigenvalues are found. The lowest eigenvalue denotes the ground state of the system and its associated eigenvector, $\mathbf{c}$, represents the coefficients of the associated superposition, i.e. $|\Psi\rangle = \sum_i c_i |i\rangle$. The energy can then be found using $E = \mathbf{c}^T \mathcal{H} \mathbf{c}$ and the forces can be computed via the Hellmann-Feynman theorem,

$$\mathbf{F}_n = -\langle \Psi| \frac{\partial H}{\partial \mathbf{x}_n} |\Psi\rangle = -\sum_{i,j} c_i c_j \frac{\partial \langle i| H |j\rangle}{\partial \mathbf{x}_n} \tag{6.1.2}$$

where $\mathbf{x}_n$ and $\mathbf{F}_n$ indicate the position and the force vector respectively for atom n. The next section provides a practical example of applying the EVB model.

6.1.1 Example case

The empirical valence bond model is based upon the creation of an adiabatic potential energy surface which is constructed by the superposition of a series of basis states. A convenient and practical choice for describing these basis states is an empirical potential such as an AMBER type potential (see Section **3.6**) however in practice this could be applied to any underlying potential that makes use of the basis states description but lacks a procedure for reactions to occur. This section will go

through a simple example of constructing such an adiabatic surface from these diabatic states to illustrate the flexibility of the EVB approach.

We will start with a simple system consisting of two basis states and construct an adiabatic surface. To do this first we first define the superposition, Ψ such that,

$$\Psi = c_1\psi_1 + c_2\psi_2 \tag{6.1.3}$$

where c_i is the coefficient for diabatic state ψ_i where $i = 1, 2$. Next we construct the usual eigenvalue equation such that,

$$\underline{\underline{H}}\mathbf{\Psi} = \varepsilon\mathbf{\Psi} \tag{6.1.4}$$

where we have changed to matrix notation. $\underline{\underline{H}}$ and ε are the Hamiltonian matrix and the energy of the system respectively. First we must find the elements of $\underline{\underline{H}}$, we do this by multiplying Eq. (**6.1.4**) by $(\mathbf{\Psi}^*)^\intercal$ and integrating over all space such that $\varepsilon = \sum_{i,j} \langle i| H |j\rangle$. The diagonal terms describe the energy of the diabatic state according to the definition of potential, this value is generally trivial to calculate. The off-diagonal terms represent the coupling between states and is typically evaluated using an empirical function based on geometry arguments involving a characteristic reaction coordinate. For the purposes of this example the Hamiltonian has been defined by quadratic functions for the classical energy of the system and a Gaussian coupling term such that,

$$\begin{aligned} H_{11} &= 2(x+2)^2 \\ H_{22} &= 4(x-2)^2 + 12 \\ H_{12} &= -14\sqrt{\frac{1}{2\pi}}\exp\left(-x^2\right) = H_{21} \end{aligned} \tag{6.1.5}$$

where x is the reaction coordinate for this system. The values in 6.1.5 were chosen as these give rise to Figure **6.2** which shows the features of the EVB methodology clearly for the purposes of this example.

Now that $\underline{\underline{H}}$ has been defined it is left to solve the eigenvalue-eigenvector equation to get two eigenvectors, Ψ_1 and Ψ_2. One of these solutions is the ground state of the system having the lowest energy and the other will represent the excited state for this system. The approach assumes that the system is only found in its ground state and so any excited states of the system are ignored.

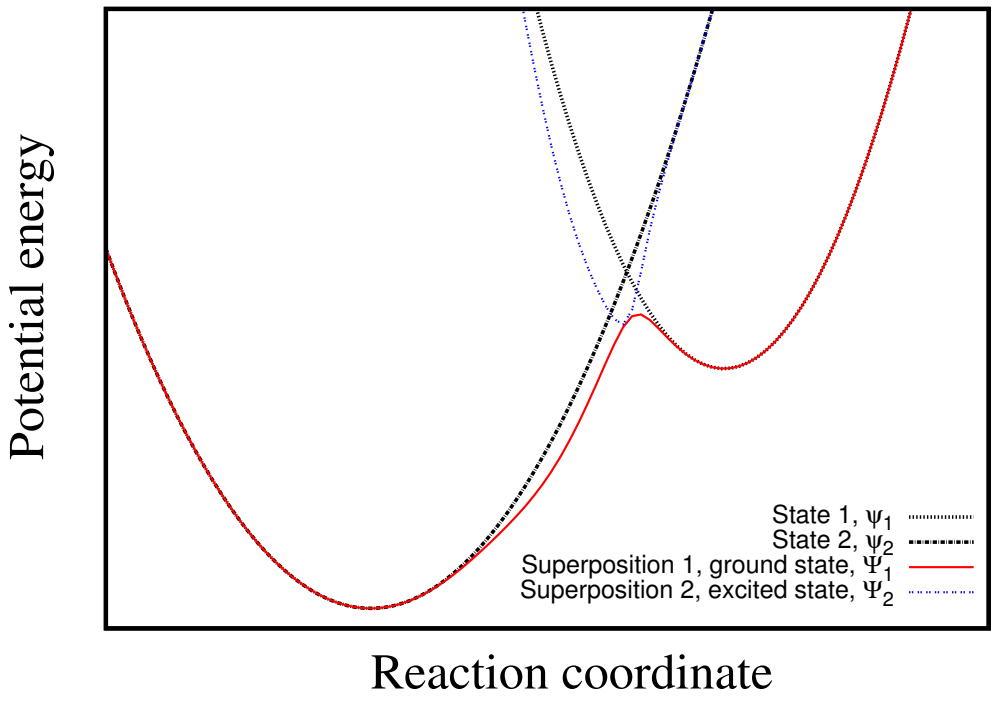


Figure 6.2: The potential energy of a two state system is shown along with the two superposition states according to Eq. (**6.1.5**).

Figure **6.2** shows the solution gained from this simple example, the superposition has lowered the barrier from the localised minimum on the right hand side and the global minimum on the left hand side[3], this has implications for the rate at which systems will traverse this barrier. The EVB methodology has created a smooth energy surface between the two basis states which is important when considering the practicalities of running molecular dynamics simulations.

This simple example shows the power of the empirical valence bond methodology to add the ability for reactions to occur within a non-reactive method. The on diagonal elements in this 2 basis state example represent the product and reactant states, and the off diagonal represents the coupling between these two states, this can be generalised to n basis states. The off-diagonal term and the number of excited states to be considered are important considerations for the method. The off-diagonal term is typically an empirical term based upon geometric arguments specific to a reaction coordinate which is characteristic of the reaction being represented by the EVB model. The number of states to be considered is ideally all of the possible states of the system with respect to the bonding, however the number also controls the size of the matrix for which the eigenvalues and eigenvectors must be found. Therefore it is computationally expensive to include all possible states,

[3]This is a well known feature of Marcus theory [**87**, **147**].

it is convenient that a majority of states will have a negligible contribution to the shape of the energy surface and so can be safely ignored.

6.2 EVB models

6.2.1 The MS-EVB3 model

In 2008 Wu *et al.* [**103**] presented the MS-EVB3 model which is designed for simulating systems with one excess proton and several water molecules. Diagonal elements of the matrix $\mathcal{H}$ are evaluated from the SPC/Fw flexible model for water [**148**] and a purpose built model for the hydronium molecule[4]. The Off-diagonal components of $\mathcal{H}$ are calculated in the following way,

$$\langle i|\, H\, |j\rangle_{i\neq j} = (V^{ij}_{const} + V^{ij}_{ex}) \cdot A(R_{OO}, \mathbf{q}), \tag{6.2.1}$$

where R_{OO} is the oxygen-oxygen distance in the hydrogen bond that includes the transferring proton, V^{ij}_{const} is a constant and V^{ij}_{ex} describes an electrostatic contribution derived from the interactions between the zundel cation that the excess proton is a constiuent of and the remaining water molecules in the system. This is given is given by,

$$V^{ij}_{ex} = \sum_{m}^{7} \sum_{k}^{N_{H_2O}-1} \sum_{n_k}^{3} \frac{q^{H_2O}_{n_k} q^{ex}_{m}}{R_{mn_k}} \tag{6.2.2}$$

where m is a label for the seven atoms that make up the $H_5O_2^+$ Zundel cation which is made of a hydronium and a water molecule between which the proton is considered to be transferring. k is a label for the remaining water molecules (N_{H_2O} is the total number of water molecules in the system) with n_k then being summed over the three constituent atoms of the water molecule. $q^{H_2O}_{n_k}$ are the partial atomic charges for the three atoms within the water molecule and q^{ex}_{m} is a set of partial charges describing the Zundel cation. R_{mn_k} is the inter-atomic distance between the atoms indicated by m and n_k respectively. The form of function $A(R_{OO}, \mathbf{q})$ was chosen to dampen unfavourable configurations and has the form,

[4]Importantly the model uses a Morse spring potential rather than a harmonic spring which gives greater flexibility in the behaviour of a proton transfer between the hydronium and a water molecule.

MS-EVB 3 off-diagonal parameters

Parameter	Value	Units
V_{const}^{ij}	-23.1871874	kcal/mol
γ	1.85	Å^{-2}
P	0.23272460	
k	9.562153	Å^{-2}
D_{OO}	2.94	Å
β	4.50	Å^{-1}
R_{OO}^{0}	3.1	Å
P'	10.8831327	
α	15.0	Å^{-1}
r_{OO}^{0}	1.8136426	Å

Table 6.1: The off-diagonal parameters of the MS-EVB3 model. Values are taken from references **[149]** and **[103]**.

$$\begin{aligned} A(R_{OO},\mathbf{q}) = \exp(-\gamma\mathbf{q}^2)\cdot\left\{1+P\exp\left[-\mathrm{k}(R_{OO}-D_{OO})^2\right]\right\} \\ \times\left\{\tfrac{1}{2}\left\{1-\tanh\left[\beta(R_{OO}-R_{OO}^{0})\right]\right\}+\right. \\ \left. P'\exp\left[-\alpha(R_{OO}-r_{OO}^{0})\right]\right\} \end{aligned} \tag{6.2.3}$$

here γ, P, k, D_{OO}, β, R_{OO}^{0}, P', α and r_{OO}^{0} are empirical parameters (see Table **6.1**) and $\mathbf{q} = \frac{1}{2}(\mathbf{r}_{O1}+\mathbf{r}_{O2})-\mathbf{r}_H$ where $\mathbf{r}_{O1}$, $\mathbf{r}_{O2}$ and $\mathbf{r}_H$ are the positions of the two oxygens and hydrogen, respectively, that participate in the hydrogen bond where proton transfer is being considered. For convenience this notation along with R_{OO} are used throughout the chapter. However the exact form of the coupling term is empirical and so it can be modified (such a modification is made in Section **6.3.4**), reference **[150]** goes further by using the genetic algorithm approach to determine the form of the off-diagonal term.

6.2.2 The SCI-MS-EVB procedure

The EVB method has the ability to simulate mixtures of several hydroniums (also known as excess protons[5]) together with water. This extension increases the size of $\mathcal{H}$ to order m^n where m is the number of states considered per excess proton (the size of the basis set for a one proton system) and

[5]A hydronium can be thought of as a water molecule bonded to a proton. It is convenient in this section to then use the phrase 'excess proton' as a label for a hydronium molecule.

n is the number of excess protons. The SCI-MS-EVB model was developed by Wang and Voth [**151**] to improve the scaling of the $\mathcal{H}$ matrix with respect to the number of excess protons. This is achieved by first treating the system as n single excess protons and performing the usual EVB methodology which results in n eigenvectors. Then each of these eigenvectors is corrected iteratively by taking into account the effects of the other excess protons. These corrections establish the environment around each excess proton. The size scaling of $\mathcal{H}$ is thus reduced from one matrix of size m^n to a consideration of n matrices of size m. The number of iterations required to reach a convergence is typically small.

6.3 EVB model for sulphuric acid and water

The following sections describe the construction of an EVB model for a system containing sulphuric acid, bisulphate, hydronium and water species. A summary of the approach is provided in Section **6.3.7**. In the discussion a naming convention is used where the `ground state` refers to the EVB state which has been identified as the most natural[6], based upon the configuration of the atoms. `Excited states` are bonding configurations that are identical to the `ground state` except for the repositioning of one bond (always involving a proton transfer). Therefore it is possible to identify an `excited state` based upon the `ground state` and the two species involved in a proton transfer. These are then referred to as a `donor` species (the molecule to which the hydrogen atom is bonded in the `ground state` and which can either be a sulphuric acid or a hydronium molecule), and an `acceptor` species (the molecule to which the hydrogen atom is bonded in the `excited state` and which can either be a bisulphate or a water molecule). For clarity the terms neutral and ionised in this chapter are used to identify the state of the sulphur-bearing species (either as a neutral sulphuric acid or an ionised bisulphate).

6.3.1 Basis set size

The basis set size for each transferable proton was chosen to be two. The reason for this choice was to keep the model as simple as possible so as to have more time to refine and use the model, in line with

[6]Natural here refers to the most logical bonding configuration given the current configuration of atoms and the possible molecules that the atoms within the system can be a constituent of (in this case this refers to sulphuric acid, bisuphate, water, and hydronium).

the SCI-MS-EVB approach implemented for water described in Section **6.2.2**. However this approach only allows each species in the system to be involved in a maximum of one proton transfer and it is therefore assumed that the possibility that a species is involved in two proton transfers simultaneously can be safely neglected. Additionally the small basis set leads to occasional ambiguities concerning which proton is considered to be transferring. A modification of the off-diagonal term as described by Eq. (**6.2.3**) has been made to limit the impact of these issues and is described in Section **6.3.4.1**. A possible extension of the model presented here would be to increase the basis state, this would require additional programming and parametrisation work.

6.3.2 Algorithms

A vital part of the EVB methodology is the construction of appropriate of basis states of the system, this is nontrivial as reactions can cause atoms to associate with different molecules throughout the simulation. This section describes two algorithms which are applied at every time step and identify the ground state and excited states of the system purely though geometric arguments. The Avogadro (version 1.0.3) **[137]** program was used to create initial configuration by geometry optimising structures taken from the literature (see Section **2.2.1**).

6.3.2.1 Ground state selector algorithm

An appropriate algorithm has been designed which constructs the `ground state` from a list of atomic positions. For convenience four molecular lists are defined and are denoted by `SA` (sulphuric acid), `BS` (bisulphate), `HY` (hydronium) and `WA` (water) and an assignment refers to the identification of an atom as a constituent of a particular molecule. The algorithm is performed in the following way:

1. The four oxygens closest to each sulphur atom are identified and assigned to a sulphuric acid molecule (along with the sulphur atom) and added to the `SA` list.

2. The remaining oxygen atoms are assigned to hydronium molecules and added to the `HY` list. The hydronium are found first for a practical consideration as it is was found easier to convert a hydronium molecule to a water molecule than vice versa.

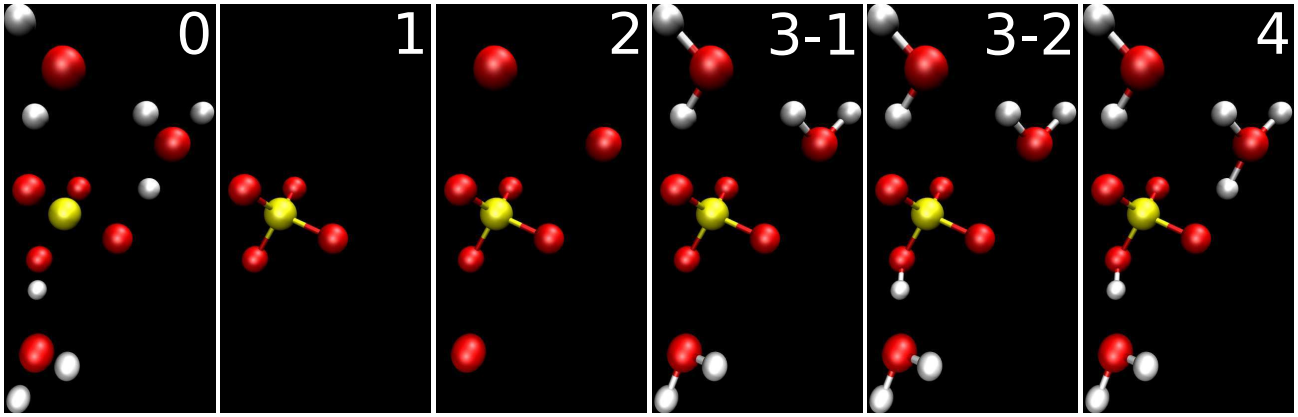


Figure 6.3: Image showing the steps of the ground state selector algorithm. Step zero shows the original positions of atoms. In subsequent steps bonds are inserted according to the procedure described in Section 6.3.2.1. Step three has been split into two parts, first to attach two hydrogen atoms to each (water) oxygen and secondly, to attach one hydrogen atom to each sulphate.

3. The minimum numbers of hydrogen atoms are assigned to each molecular species. For each member of the `HY` list the two hydrogen atoms closest to the associated oxygen are identified and assigned to the list. The same process is used to assign one hydrogen to each member of the `SA` list.

4. The remaining hydrogen atoms are placed in a temporary list named `H` (in no particular order). The closest oxygen to each of these hydrogens is identified and the hydrogen is assigned to the molecular species of which the oxygen is a constituent. The oxygen cannot be part of a molecule which has already accepted one of these hydrogens and cannot be an oxygen which is a member of the `SA` list which has been bonded to a hydrogen atom in step 3. If there is an attempt to assign a hydrogen where the bond length is greater than 1.2Å then this assignment is rejected and the atom is moved to the top of the `H` list. If the `H` list is, at any time, rearranged then step 4 is immediately restarted using the modified `H` list. The hydrogens which have been previously rearranged in the `H` list are no longer subject to the 1.2Å constraint; but if the bond length is over 2Å the hydrogen is again returned to the top of the `H` list. If problems with assignments persist then the algorithm is run with the bond length check turned off. In practice there is only need for these checks in the circumstances of proton transfer events, in which case the two-state EVB mixing between the uncertain states will ensure that the appropriate forces are applied regardless of the assignment of the molecular species. This step is completed when all hydrogens in the `H` list have been assigned. The criterion applied in this step were found

through trial and error experiments with simulation of the model.

5. The previous step ensures that all hydrogen atoms have now been attached to a molecule (i.e. there are now no atoms waiting to be assigned). Therefore any sulphuric acid molecule in the `SA` list where only one hydrogen atom has been assigned is moved to the `BS` list. Similarly any hydronium molecules in the `HY` list that do not have three assigned hydrogen atoms are moved to the `WA` list.

Figure **6.3** illustrates the steps of the algorithm as they are performed on a system. The `SA`, `BS`, `HY` and `WA` arrays now represent the `ground state` of the system based upon geometric arguments.

6.3.2.2 Excited states identification algorithm

As stated earlier, an `excited state` is defined as a change in the bonding assignment of one hydrogen in the `ground state`. The algorithm identifies the `excited states` in a system using the following steps:

1. The shortest distance between a hydrogen belonging to a `donor` species, and an oxygen belonging to an `acceptor` species is identified. The oxygen in a bisulphate ion which has an attached hydrogen is not considered here. If the distance is less than 2Å then an `excited state` is identified where the hydrogen is reassigned to the `acceptor` species.

2. Step 1 is repeated until either the shortest distance identified is greater than 2Å in length or there are no further `acceptor` or `donor` molecules from which to construct an $O-H$ separation distance.

The ground state selector and the excited state identification algorithms construct one `ground state` and n `excited states` where n is the number of proton transfers under consideration. When $n = 0$ the system is in a non-reactive configuration. Figure **6.4** shows a configuration in the ground state of the system, there are then three possible excited states of the system (which have been labelled), thus the `ground state` and three `excited states` have been identified in this system.

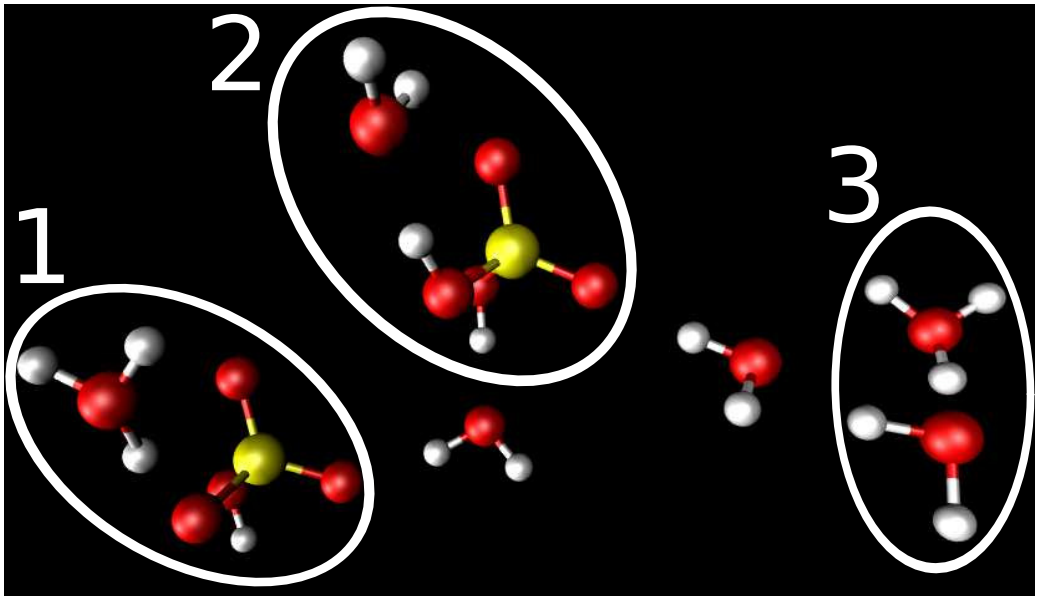


Figure 6.4: Image showing a system with three separate proton transfer events being considered. They have been labelled so that groups 1, 2 and 3 have the potential for a [hydronium/bisulphate], [sulphuric acid/water] and a [hydronium/water] proton transfer respectively. One `ground state` and three `excited states` have therefore been identified.

6.3.3 Diagonal terms

The choice of potential for the diagonal term was made on a practical consideration as the majority of the effort was used in developing the off-diagonal potential. The potentials used to specify the diagonal elements of $\mathcal{H}$ are those provided by Loukonen *et al.* [**152**], the SPC/EF potential is used for the water molecules [**153**]. The hydronium however is represented by a harmonic angle potential and a Morse bond potential so as to match that of the MS-EVB3 model, it has the form of,

$$V_{hydonium} = \frac{1}{2}\sum_{i=1}^{3} k_\theta \left(\theta_i - \theta_0\right)^2 + \sum_{i=1}^{3} D\left[1 - e^{-\alpha(r_i - r_0)}\right]^2 \tag{6.3.1}$$

where i is summed over the three hydrogen-oxygen-hydrogen angle bonds and the three oxygen-hydrogen bonds respectively. The angle parameters are taken from reference [**152**]. The Morse potential is taken from reference [**103**], however the α parameter was changed such that the second order Taylor expansion around r_0 matched the harmonic spring provided by reference [**152**], and so $\alpha = 2.327\text{Å}^{-1}$. This was done as the parametrisation work was performed with the harmonic bonds, however the Morse potential was found to improve the match between this work and that of the MS-EVB3 model.

The diagonal term is then augmented by an energy shift, Δ, to account for the difference in

zero temperature ground state energy between the neutral pair (sulphuric acid and water) and an ionised pair (bisulphate and hydronium). Ding *et al.* [**96**] calculated the value of this parameter to be $\Delta = 144.0\,\text{kcal/mol}$, but for the purposes of this model it was considered to be a free parameter[7] which is to be fitted in Chapter **7**.

6.3.4 Modifications made to the off-diagonal term

The empirical form of the off-diagonal term is based upon the MS-EVB3 model (Eq. (**6.2.3**)), with two modifications. These relate to the expression involving the $\mathbf{q}$ parameter (Section **6.3.4.1**) and the method for calculating V_{ex}^{ij} (Section **6.3.4.2**).

6.3.4.1 The q dependence

The dependence of the off-diagonal term on $\mathbf{q}$ has been modified to reduce the possibility of instabilities caused by limiting the number of EVB basis states to two per proton. The two basis states led to discontinuities when switching the molecules involved in a proton transfer and this caused issues in simulations. The term in Eq. (**6.2.3**) involving $\mathbf{q}$ was thus modified as follows,

$$\exp(-\gamma\mathbf{q}^2) \rightarrow \begin{cases} \left[\frac{\exp(-\gamma\mathbf{q}_0^2)-\exp(-\gamma)}{1-\exp(-\gamma)}\right] & \mathbf{q}_0^2 \leq 1 \\ 0 & \mathbf{q}_0^2 > 1 \end{cases} \tag{6.3.2}$$

where $\mathbf{q}_0^2$ is related to $\mathbf{q}$ in the following way. The $\mathbf{q}$ vector is transformed ($\mathbf{q} \rightarrow \mathbf{q}'$) to a new orthogonal coordinate system in which the x-axis is parallel to $\mathbf{r}_{OO}$ where $\mathbf{r}_{OO} = \mathbf{r}_{O1} - \mathbf{r}_{O2}$. The other two axes have an arbitrary direction. $\mathbf{q}_0^2$ is then defined as $\mathbf{q}_0^2 = \left(\frac{q'_x}{\tau_x}\right)^2 + \left(\frac{q'_y}{\tau_{yz}}\right)^2 + \left(\frac{q'_z}{\tau_{yz}}\right)^2$ where τ_x and τ_{yz} correspond to the semi-major axes of an ellipsoid with respect to the x axis, and the y and z axes respectively. In effect, an ellipsoidal surface replaces that of the sphere in the MS-EVB3 method. This allows us to restrict the available range of hydrogen positions along the $\hat{\mathbf{r}}_{OO}$ direction. The form of Eq. (**6.3.2**) ensures two features: that the factor is equal to 1 at $\mathbf{q}_0^2 = 0$ and goes to zero at $\mathbf{q}_0^2 = 1$. For $\mathbf{q}_0^2 > 1$ the off-diagonal term is set to zero which enforces a classical behaviour outside the ellipsoid defined by $\mathbf{q}_0^2 = 1$.

The benefit of this modification is the ability to treat each axis in the ellipsoid differently, i.e

[7] This allows for greater flexibility in parametrising the off-diagonal term and improve the fit.

use a shorter maximum range for the x axis which reduces the likelihood of switching behaviour. This is where the `donor` molecule switches which molecule is acting as the `acceptor` molecule in one time step which causes instabilities in the potential energy landscape. Two parameters are introduced in this modification, τ_x and τ_{yz}, these are treated as additional parameters to fit.

6.3.4.2 Intermediate charges

The V_{ex}^{ij} parameter as defined in Eq. (**6.2.2**) is also modified as the charges provided by the MS-EVB3 model for the calculation of this term are specific to the Zundel cation and are not applicable for a sulphuric acid and water proton transfer. An alternative approach is to determine the charges of the atoms in the two molecules, between which the proton is considered to be transferring, according to their c_i values. The approach is as follows:

1. An initial guess for the charges are made by assuming $q = 0.5(q^{ground} + q^{excited})$ where q^{ground} and $q^{excited}$ are the partial charges for that atom according to the classical potential for the molecule, where the molecule is identified by the `ground state` and the `excited state` respectively.

2. A self-consistent iteration is performed where q is updated at each iteration according to $q = c_{ground}^2 q^{ground} + c_{excited}^2 q^{excited}$ where c_{ground} and $c_{excited}$ are the **c** eigenvector coefficients for the `ground state` and `excited state` respectively.

This procedure is performed at every time step for [sulphuric acid/water] or [bisulphate/hydronium] proton transfers alongside a separate self-consistent iterative process described in the next section which has a purpose to allow consideration of multiple proton transfers. For [water/hydronium] proton transfers the zundel charges as described in the MS-EVB3 model were used.

6.3.5 Multiple-proton procedure

The model presented here follows the SCI-MS-EVB procedure developed by Wang and Voth [**151**]. Figure **6.4** shows a system where this extension is required as it has three `excited states`. The initial guess in this procedure neglects overlap between two different `excited states` (i.e. $\langle i| \mathcal{H} |j\rangle = 0$ where i and j represent two different `excited states`). This allows for the deconstruction of the EVB Hamilton of the system ($\mathcal{H}$) which is of size $n_d^{n_c}$ to n_d matrices of size n_c where n_d is the

number of `donor` molecules being considered and n_c is the number basis states considered per `donor` molecule. An iterative procedure is then used to correct these matrices for $\langle i| \mathcal{H} |j\rangle$ contributions. The method can be described as follows:

1. Calculate the eigenvector for each 2×2 matrix $\mathcal{H}$ defined by the `ground state` and an `excited state`. For convenience n is used as a label such that $\mathbf{c}_n$ and $\mathcal{H}_n$ are defined for `excited state` n. In addition n is used to label the two molecules which make up the `donor` and `acceptor` species which define the `excited state`.

2. Each $\mathcal{H}_n$ is then corrected for `excited state` m where $m \neq n$. This correction is for the intermolecular interactions between the `acceptor/donor` pairs described by n and m. This is performed by updating the intermolecular energy contributions to the diagonal term in $\mathcal{H}_n$, i.e. $E^{ground}_{intermolecular} \rightarrow \left(c^2_{ground} E^{ground}_{intermolecular} + c^2_{excited} E^{excited}_{intermolecular}\right)$ where c^2_{ground} and $c^2_{excited}$ are the squared coefficients of the $\mathbf{c}_m$ eigenvector which represent the associated `ground state` and `excited state` respectively. $E^{ground}_{intermolecular}$ and $E^{excited}_{intermolecular}$ are the intermolecular energy contributions resulting from interactions between the `acceptor/donor` pair n and the `acceptor/donor` pair m in the `ground state` and `excited state` of m respectively. The term V^{ij}_{ex} in the off-diagonal expression is updated in the same fashion; this includes recalculating the charges q according to Section **6.3.4.2**. The $\mathbf{c}_n$ values are then recalculated. This allows the coefficients $\mathbf{c}_n$ for the `acceptor/donor` pair n to be corrected for the intermolecular energy contributions from `acceptor/donor` pair m. This step is repeated until each $\mathbf{c}_n$ vector has been corrected for each `acceptor/donor` pair m (where $m \neq n$).

3. Step 2 is repeated until all $\mathbf{c}_n$ eigenvectors have converged. The convergence is tested by defining $c_{sum} = \sum_n (\mathbf{c}^{old}_n - \mathbf{c}^{new}_n)^2$ where $\mathbf{c}^{old}_n$ and $\mathbf{c}^{new}_n$ are the $\mathbf{c}_n$ values calculated before and after step 2 is performed respectively. The system is considered to be converged when $c_{sum} < 10^{-5}$, or when step 2 has been cycled 10 times. In practice only a few iterations are required, the hard limit in the total number of steps was a practical consideration which ensured that the simulation could not get stuck in an infinite loop.

6.3.6 Energy and force calculation

Once the self-consistency iterations have been performed the energy of the system can be computed in the same fashion as in the SCI-MS-EVB method. For example Eq. (**13**) in reference [**151**] gives a decomposition of the total energy of a system containing two excess protons in the form

$$E_{total} = E_{AA} + E_{BB} + E_{AB} + E_{AR} + E_{BR} + E_{RR} \tag{6.3.3}$$

where E_{total} refers to the total energy of a system according to the EVB method. A and B refer to separate `acceptor/donor` pairs (see Figure 1 in reference [**151**]). E_{AA}/E_{BB} refer to the independent energy contributions of the `acceptor/donor` pairs and includes their off-diagonal contributions. E_{AB} describes the energy contribution due to interactions between A and B. R refers to the rest of the system, and there are three further contributions due to interactions between A and R, B and R and the independent energy contribution of R, referred to as E_{AR}, E_{BR} and E_{RR} respectively. Eq. (**6.3.3**) can be generalised to $E_{total} = E_{RR} + \sum_i \left(E_{i,i} + E_{i,R} + \sum_j^{i \neq j} E_{i,j} \right)$ where i and j refer to identified `donor/acceptor` pairs. Once E_{total} has been constructed it simply remains to apply the Hellmann-Feynman theorem to determine the force acting upon each atom within the system.

6.3.7 Model Overview

In summary, the model can be described as an algorithm consisting of the following series of steps:

Step 1 Construct a `ground state` from a set of atomic positions (Section **6.3.2.1**).

Step 2 Identify the `excited states` (Section **6.3.2.2**).

Step 3 Calculate the on-diagonal and off-diagonal terms of the 2×2 matrix $\mathcal{H}$ for each `excited state` (Section **6.3.3** and **6.3.4.1**).

Step 4 Optimise the $\mathbf{c}_i$ vectors for each `excited state` by performing a self-consistent iterative procedure which revises the off-diagonal terms and allows multiple proton transfer to be accommodated (Section **6.3.4.2** and **6.3.5**).

Step 5 Calculate the energy of the system and the forces acting upon each atom (Section **6.3.6**).

6.3.8 DL _POLY extension

During the course of the project several modules of FORTRAN 90/95 code was written and integrated into DL_POLY 4.0.3. This third party extension of DL_POLY contains 16,692 lines of code and therefore it is not feasible to attach the code to the thesis. However this section includes a brief overview of the code.

In FORTRAN 90/95 object oriented code is made possible through the use of types. Table **6.2** describes the types which were defined for the EVB code and which were used extensively as a framework to perform the necessary calculations for implementing the EVB methodology. There are also several subroutines and functions supplied for performing calculations upon these objects.

Figure **6.5** shows a schematic of the EVB code. The approach taken was to make the minimal changes necessary to the DL_POLY code and to use the original source for testing purposes. Therefore it was decided to group all the methods within DL_POLY and then use a keyword within the CONTROL file to switch between using the DL_POLY calculations and using the EVB code, this will construct the molecules from the atomic list supplied by the DL_POLY code and will then calculate energies and forces. A second switch was used to activate the EVB adiabatic surface over that of a simple classical picture.

The run time for a serial 1 ns simulation with six water molecules and one sulphuric acid was 430 seconds on an Intel Core i5-460M processor and it is therefore easily within reach to simulate tens of nanoseconds of behaviour using this model. This is within the reach of the project's goal of simulating small clusters of sulphuric acid and water to construct a free energy landscape via the use of thermodynamic integration, this free energy landscape would then yield a nucleation rate.

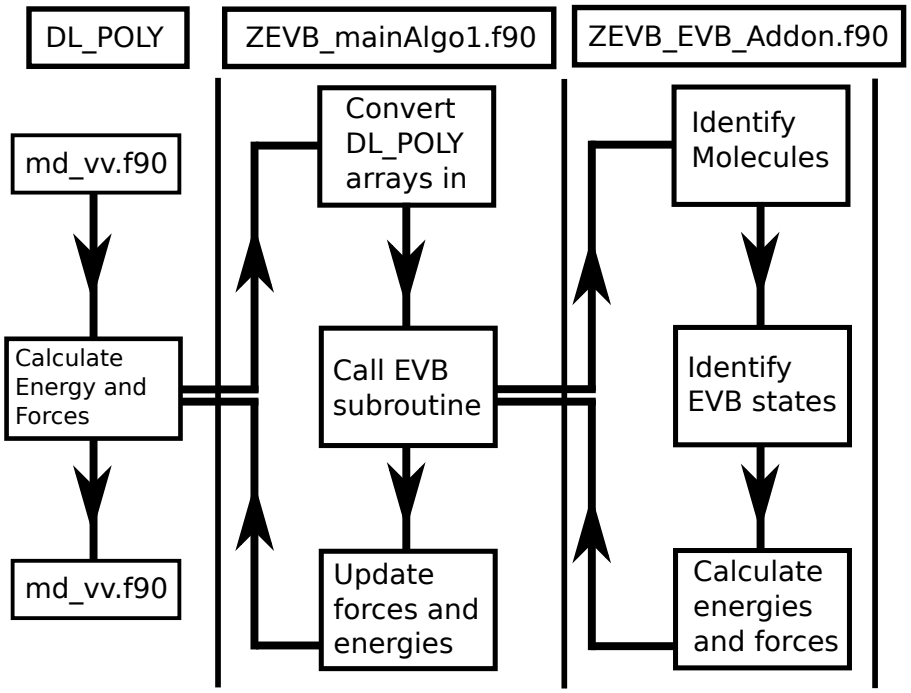


Figure 6.5: Schematic of the EVB code.

Name	Contains	Description
Atom	Position vector (real), force vector (real), species name (character), place (integer, for integration with DL_POLY)	All essential information required when quering an atom.
Water	Three atoms	Contains all atoms assigned to the water molecule.
Hydronium	Four atoms	Contains all atoms assigned to the hydronium molecule.
Bisulphate	Six atoms	Contains all atoms assigned to the bisulphate molecule.
Sulphuric acid	Seven atoms	Contains all atoms assigned to the sulphuric acid molecule.
Stater	Four allocated lists, one for each of the molecules defined above.	A possible system state containing molecules of four types.
EVBStater	Two staters. It also contains various bits of information important in the evaluation of the EVB method.	Detailed information important for performing the EVB method.

Table 6.2: Details of the defined types within the FORTRAN 90/95 code written for this project.

Chapter 7

Parametrisation of the EVB model

The EVB model developed in the previous section is based upon the MS-EVB3 model [**103**] which was designed for use with water and one hydronium in the bulk. However for this study we are instead interested in clusters of sulphuric acid and water which raises the question of the validity of the parameters used in the MS-EVB3 model. The most serious of these questions is the correct parameters to use for the [sulphuric acid/water] proton transfer. The literature (see Chapter **2**) suggests that monohydrated or dihydrated suphuric acid will remain neutral whereas tetrahydrated and higher hydrated clusters will be ionised, though there is some question around the trihydrated case. The recreation of this behaviour using the EVB methodology is a key goal for the model. To this end it was decided that the EVB model describing the [sulphuric acid/water] proton transfer would be parametrised such that it matched the higher level theory.

The [sulphuric acid/bisulphate] proton transfer was considered to be less important and so used the same parameters found for the [sulphuric acid/water] proton transfer in this section. The [water/hydronium] proton transfer used the same parameters as the MS-EVB3 model (see Table **6.1**) except for R^0_{OO}, P and D_{OO}. These were modified by hand to increase the number of observed proton transfer events such that, $R^0_{OO} = 2.7$ Å, $P = 0.4$ and $D_{OO} = 2.65$ Å. This modification was performed by visual analysis of simulation trajectories after the observation was made that the original set of parameters did not lead to observed [water/hydronium] proton transfer events[1]. If more time was available a suitable DFT simulation would be performed to provide data with which to fit these parameters.

[1] A possible cause of the reduced observed proton transfers in this model and that of the MS-EVB3 model is the maximum basis state limit of two per `donor` species in the system.

7.1 Particle swarm optimisation

7.1.1 Introduction

The particle swarm optimisation (PSO) [**155, 156**] was chosen as the parametrisation scheme to follow in this project due to its scalability as there were many parameters to be examined with the approach. The PSO technique mirrors molecular dynamics annealing in that it involves several particles exploring a solution space in an effort to relax to the best solution. The method makes few assumptions about the model being optimised. It is a metaheuristic method as it does not guarantee an optimal solution, this aspect of the method should be considered for applications. In the case of the parametrisation of the EVB method it is thought that there is no perfect solution, and so the use of the method to find an appropriate solution is acceptable and has the benefit of being simple to implement.

The PSO technique has several similarities with the evolutionary algorithms (EA) [**157**] group of methods. The mechanics of the swarm is controlled by a simple equation of motion which results in interesting emergent behaviour as information is shared within the swarm. This simplicity allows for many modifications to be made in attempts to improve aspects of the method. A test case was examined for the purpose of becoming familiar with the method and for writing and refining a code to perform the PSO technique upon a model.

7.1.2 Literature review

The strength of the PSO method lies in the small number of assumptions the method makes and its relatively cheap computation cost which makes it suitable when there is a large parameter space to be searched and limited data for which it is being optimised. The modern version of the model was introduced by Kennedy and Eberhart [**155**] in 1995 and was extended by Shi and Eberhart [**156**] in 1998 who added a friction term. The PSO technique is reviewed by Banks *et al.* [**158**, **159**]. The application of the PSO method has been analysed by Poli [**154**] who found that it has been used in diverse fields from engineering and design to physics and security. Figure **7.1** displays the number of papers published which made use of the PSO method over the last 20 years showing an exponential growth over $1998 - 2006$ leading to a steady use over the following 8 years.

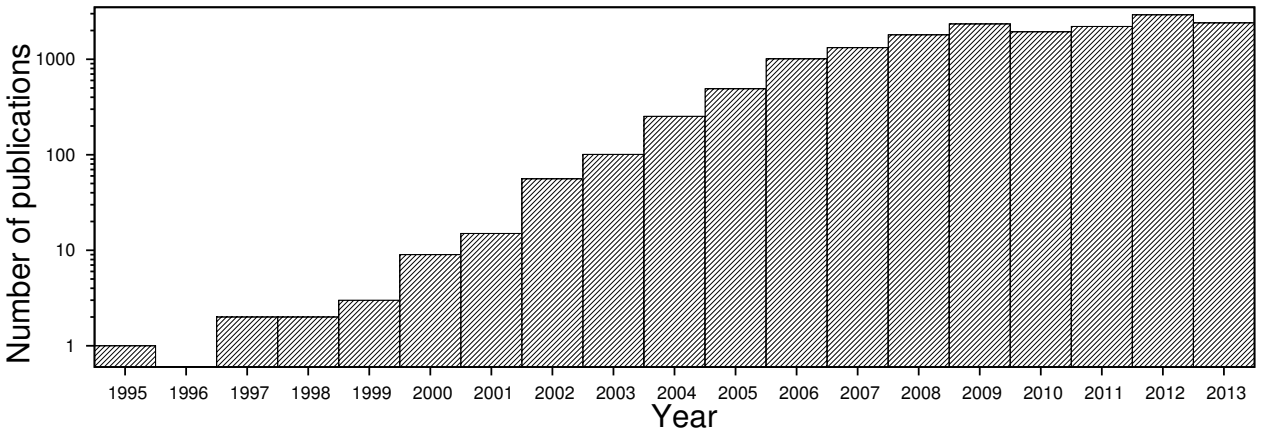


Figure 7.1: The number of publications per year found using the web of science website under the search term of 'particle swarm optimisation' and 'particle swarm optimization'. The image is comparable to Figure 4 in reference [**154**].

7.1.3 Method

The method involves simulating a swarm of particles, each particle represents a certain choice of parameters which are being fitted. These particles then move though parameter space according to an equation of motion in an analogous way to molecular dynamics. Information is shared between particles via the equation of motion and this gives rise to swarm behaviour of the particles. Each particle is influenced by its best position and the swarm's best position (as determined by a fitting function), these position are updated as necessary throughout the simulation.

PSO was first described in its modern form by Kennedy and Eberhart [**155**] with the following equation of motion,

$$v_{t+1,i} = v_{t,i} + \varphi_1\beta_1(p_i - x_{t,i}) + \varphi_2\beta_2(p_g - x_{t,i}) \tag{7.1.1}$$

$$x_{t+1,i} = x_{t,i} + v_{t+1,i} \tag{7.1.2}$$

where $x_{t,i}$ and $v_{t,i}$ represent the position and velocity respectively of particle i in parameter space at iteration t. φ_1 and φ_2 represent the influence of the local and group (swarm) respectively. β_1 and β_2 are uniformly distributed random numbers in the range $0 \leq \beta_{1,2} \leq 1$. p_i represents the best position found so far based upon the history of particle i. Here the best position is identified by a test of the particle's current location $x_{t,i}$ and is typically a comparison against reference data. p_g represents the best value found so far based upon the history of the entire swarm. This value is the social element of the procedure where information is shared between particles. Here t refers to iteration number. It

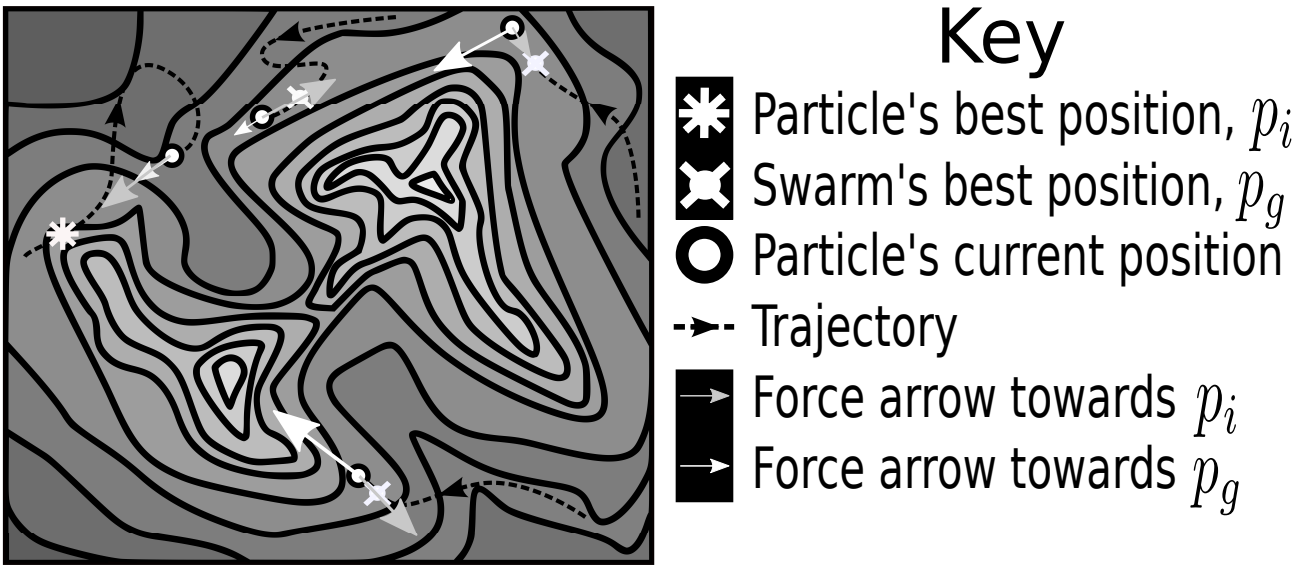


Figure 7.2: A four 'particle' PSO attempt to optimise two parameters of a model. The parameter space is represented as a 2D contour plot where the x and y directions represent the two parameters being fitted and the surface represents the fit. The dark grey areas of the surface represent a bad fit and light grey areas of the surface represent a good fit (the quality of the fit is determined by Eq. (**7.2.2**)). The particles follow trajectories across parameter space influenced by 'forces' directed towards individual and collective best fit locations. The varying length of the force arrows represents the noise introduced into the swarm through parameters β_1 and β_2.

is trivial to expand equations (**7.1.1**) and (**7.1.2**) to include multiple swarms, multiple parameters to be optimised or multiple runs. Additionally the technique is simple to code as a parallel program, increasing the efficiency of the technique[2].

A small modification of Eq. (**7.1.1**) was introduced by Shi and Eberhart [**156**] for the purposes of improving the convergence of the PSO method such that,

$$v_{t+1,i} = v_{t,i}\omega + \varphi_1\beta_1(p_i - x_{t,i}) + \varphi_2\beta_2(p_g - x_{t,i}) \tag{7.1.3}$$

where ω is a factor introduced to scale the velocity memory component of the velocity update. Scaling down of the ω parameter strengthens the attraction towards p_i and p_g and so improves the convergence towards these points.

The Fortran 90/95 programming language was used to construct a code to perform the PSO method. Care was taken to separate the model being parametrised and the PSO code such that it could be used with the EVB model.

[2]The implementation detailed in this chapter did not take advantage of parallel programming.

7.2 Test case

A test case for the PSO technique was devised to gain experience implementing the PSO technique. Eq. (**7.2.1**) was chosen as the model (here x is an independent parameter of the system),

$$f(x;a,b) = \frac{a}{10}\cos(2x+5a) + \frac{b}{10}\sin(3x+5b) \tag{7.2.1}$$

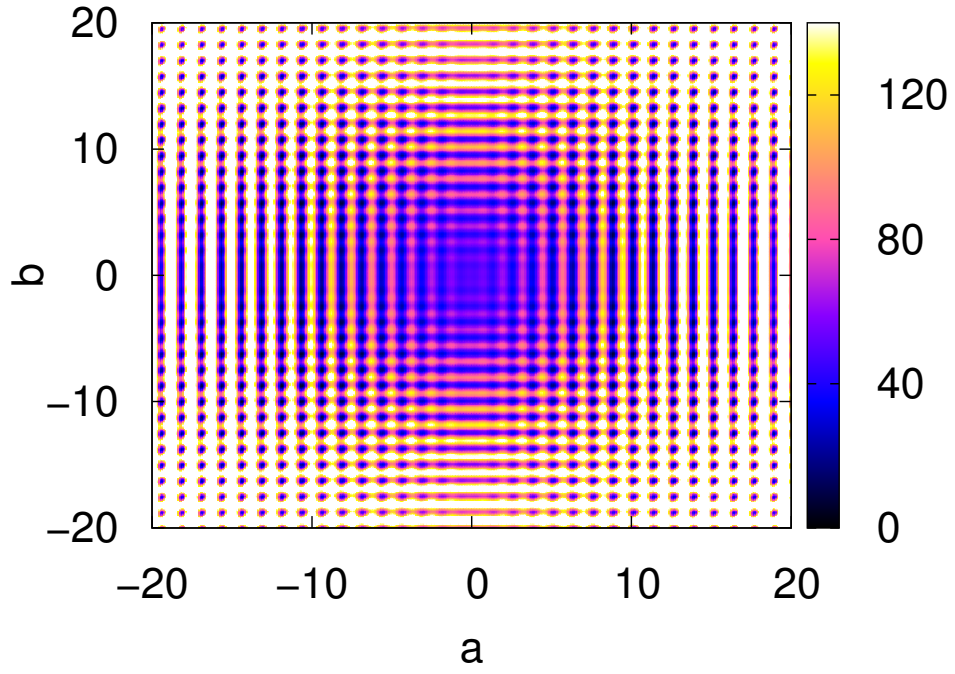
where a and b are two parameters which are to be fitted. The equation has several local minima due to its strong phase dependence and weak amplitude dependence. The task is to find the best estimate of a and b starting with data generated over a sample of x points with a given choice of a and b. The data set, Ξ, used for the fitting was created by using $a = 12.65$ and $b = 8.25$ and x in the range of 0 and 5 in steps of 0.1 (totalling 51 data points). Eq. (**7.2.1**) was used as the model and the fitting function used to assess the quality of the solution is described by,

$$Fit(a,b) = \sum_x (\Xi_x - f(x;a,b))^2 \tag{7.2.2}$$

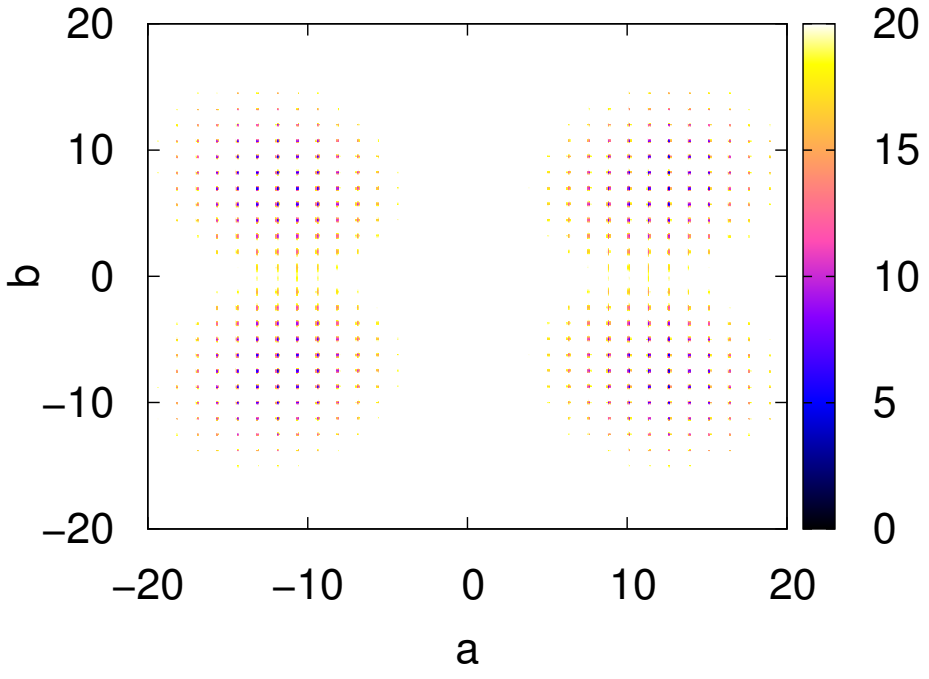
where Ξ_x refers to the reference data for the specific x. Figure **7.3a** shows a 2D contour plot of Eq. (**7.2.2**) over parameters a and b. The figure shows several local minima due to the phase in Eq. (**7.2.1**) repeating every $\frac{2}{5}\pi$ in a and b. The depth of each local minimum is different, this feature is not clear from Figure **7.3a** and so Figure **7.3b** is the same plot with a limited range to make this visible. It is now possible to see that there are four regions where the $Fit(a,b)$ is achieving values nearing zero. The four regions can be explained by separating Eq. (**7.2.1**) into an amplitude and phase part. The phase part is responsible for the large number of local minima, whereas the amplitude part only produces minima near $|a| = 12.65$ and $|b| = 8.25$. The combination of these two effects gives rise to the pattern seen in Figure **7.3**.

7.2.1 Assessment

Several features of the PSO method are assessed in this section. A histogram of fitting function (**7.2.2**) value obtained for 100 runs of the PSO method for each choice of settings is constructed and compared to assess the difference in performance given by the choice of these settings.



(a)



(b)

Figure 7.3: (A) and (B) both show the surface plot of Eq. (**7.2.2**). (A) shows the many local minima found while varying parameters a and b. More features of the surface are found by reducing the range such as in (B) where it is now clear that there are four regions where the surface is being minimised, however there is only one perfect solution where the surface goes to zero.

7.2.1.1 Particle and swarm numbers

The number of particles per swarm and the number of swarms are open questions when implementing the PSO technique. When there is only one particle per swarm the technique is essentially a non-interacting particle simulation. On the other extreme, when there is one swarm it is possible that the entire swarm becomes trapped in a local minimum. Figure **7.4** compares differing numbers of swarms, S and particles per swarm, P, while keeping the total number of particles constant.

Figure **7.4** indicates that having a small number of particles per swarm (a small P) does not perform as well as attempts made with a larger P. This could be a result of the limited information exchanged reducing the swarm behaviour of the system. The other extreme performs better indicating that it is better to choose a larger P rather than a larger S. However the best results[3] are given when (S=10, P=10) and (S=4, P=25). These numbers indicate that for this system the best result can be found by allowing there to be multiple swarms, with each swarm having a number of particles that can take advantage of the swarm behaviour of the PSO technique. This result may be applicable for systems where there are thought to be many local minima in the parameter space being searched, however this should be checked to ensure the appropriate approach is being used.

7.2.1.2 ω scaling

The scaling of the ω value can improve convergence by reducing the particle's memory of its velocity. Here ω is function of t such that $\omega(t) = \omega_{max} - (\omega_{max} - \omega_{min}) * t/t_{total}$ where ω_{max}, ω_{min} are the maximum and minimum values for ω and t_{total} is the total number of iterations in the simulation. Figure **7.5** compares different ω scaling regimes.

The conclusion drawn from Figure **7.5** is that scaling does improve the performance, however only when scaling reduces ω from ~ 0.9 to ~ 0.4. This is in agreement with other assessments of ω scaling [**158**, **159**].

7.2.1.3 Initial particle positions

The initial position of each particle is determined at random within a starting range. This test looks at how the size of this range affects the performance of the PSO technique (see Figure **7.6**).

[3]The best result here refers to achieving a low fitting value, in this case the (S=10, P=10) and (S=4, P=25) have a higher chance of returning a smaller fitting value.

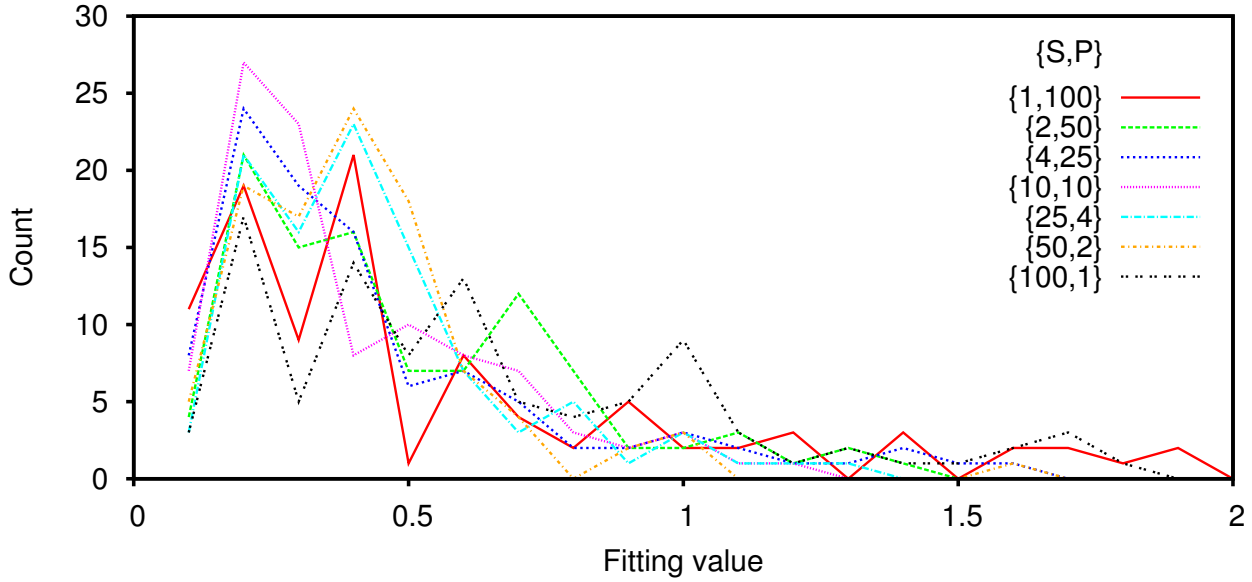


Figure 7.4: Histogram of the performance of the PSO scheme with differing number of swarms, S, and number of particles per swarm, P. The total number of particles in the simulations is a keep constant equal to 100. (see Section **7.2.1.1**)

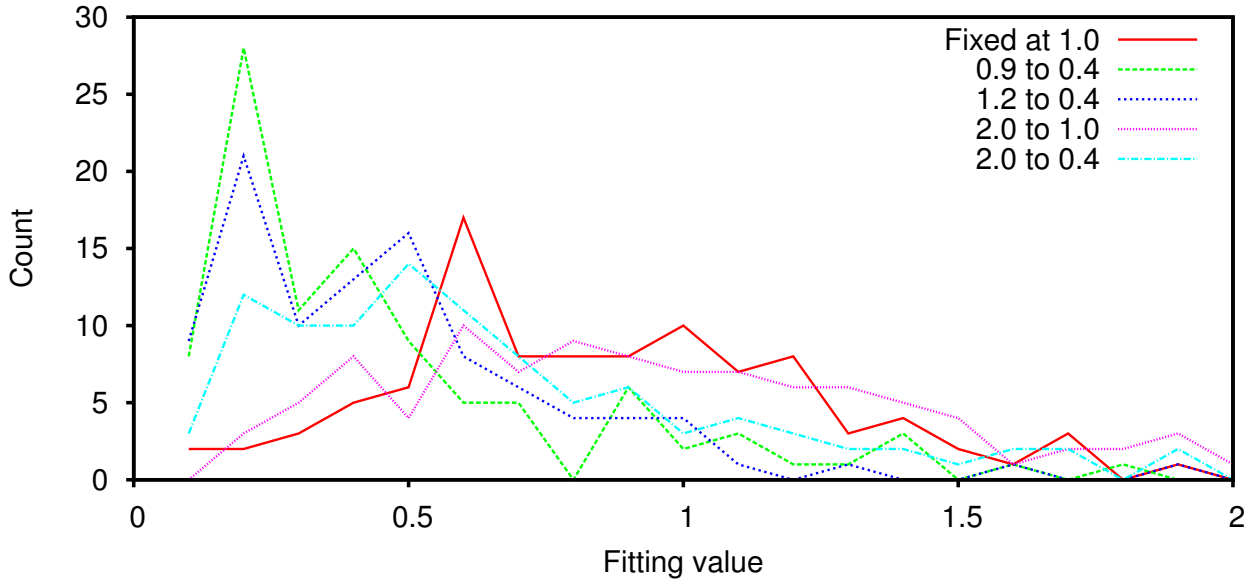


Figure 7.5: Histogram of the performance of the PSO scheme whilst using different ω scaling regimes. ω is scaled linearly over the course of the simulation from the higher value to the lower value as indicated in the legend. The fixed value does not alter the value of ω over the course of the simulations. (see Section **7.2.1.2**)

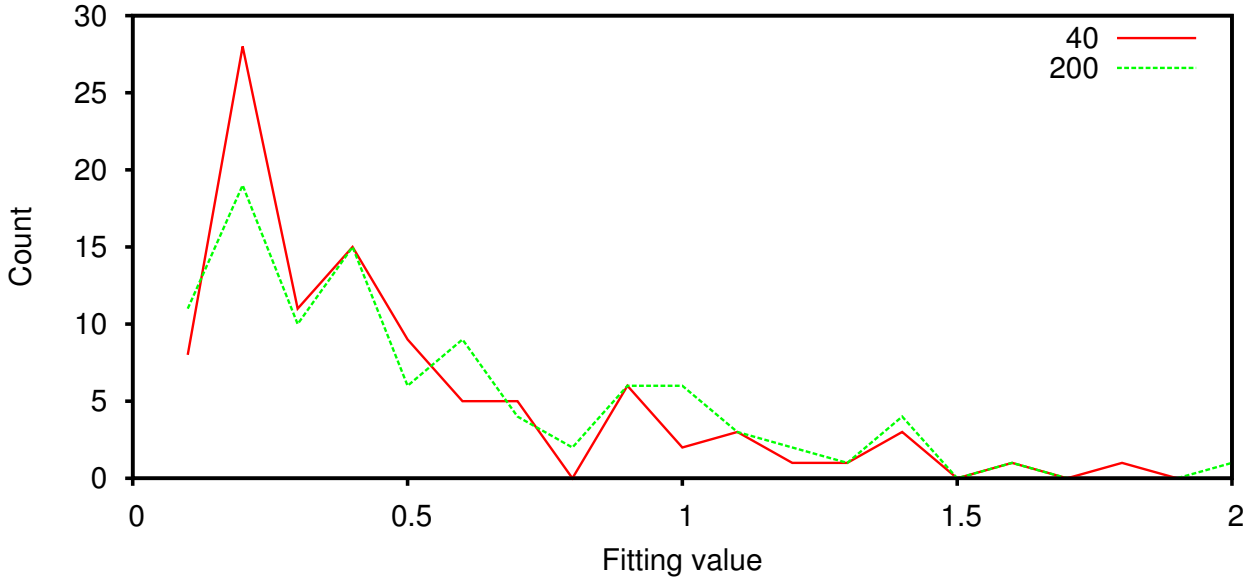


Figure 7.6: Histogram of the performance of the PSO scheme whilst changing the range in which the particles are initiated. The range referred to here is a confining range for which a parameter can be initialised and move within i.e. for the parameter κ in the range $\kappa_{min} - \kappa_{max}$, the found solution ($\kappa_{fitted}$) for the parameter will be confined such that $\kappa_{min} \leq \kappa_{fitted} \leq \kappa_{max}$. The legend represents the range of the initial values for parameters a and b, the range is centred around zero. The range given in a real system is typically set by hand through practical considerations. (see Section **7.2.1.3**)

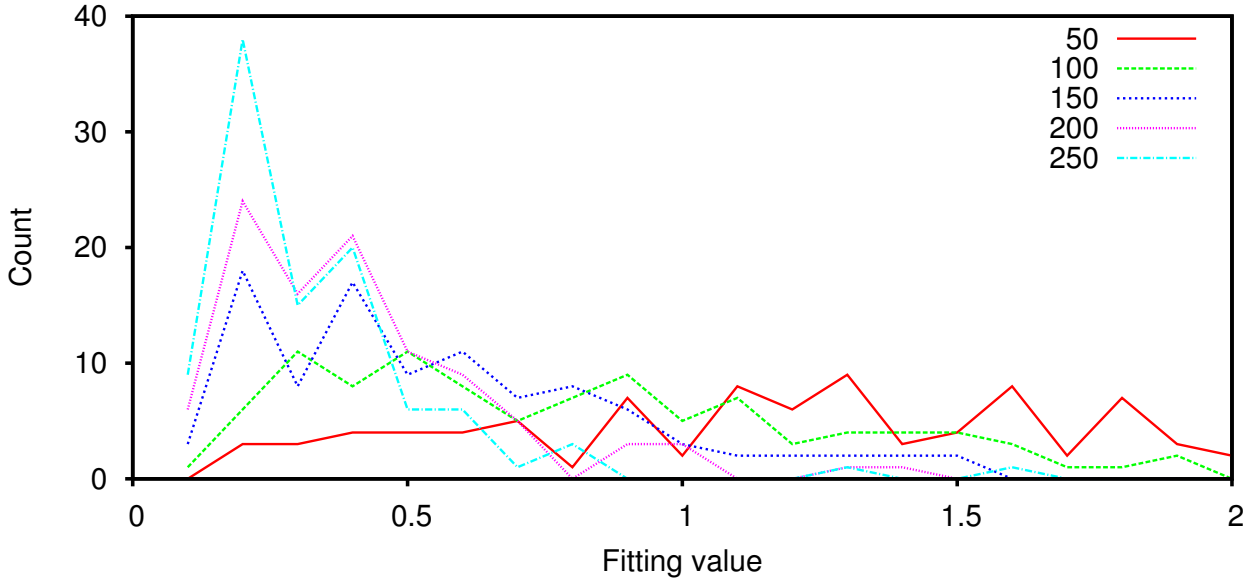


Figure 7.7: Histogram of the performance of the PSO scheme with a varying number of iterations. The legend shows the number of iterations allowed in the simulations. (see Section **7.2.1.4**)

There is a performance cost in expanding the starting range which is not surprising due to the larger range being less likely to initiate the particles close to the best solution, however the effect is small. Confining particles in the PSO technique to a certain range can help improve the efficiency of the technique and enforce constraints upon the solution.

7.2.1.4 Number of iterations

The number of iterations per PSO run is an important feature to explore. Too few and the search does not take advantage of the swarm behaviour, too many iterations would not be an efficient use of computer time. Figure **7.7** shows a histogram comparing differing total number of iterations.

From Figure **7.7** it is clear that the larger number of iterations correlates with a better performance of the PSO technique. However the optimisation performs well at 150-200 iterations. It should also be noted that this is a worst case situation with several local minima and that it is expected that fewer iterations would be needed if the fitting function is smoother than the case tested here.

7.2.1.5 Maximum velocity

The maximum velocity (where velocity is defined by change in the value of a parameter over one iteration) achievable by a particle within the PSO technique is typically limited to ensure the stability of simulations. Figure **7.8** compares the performance of the PSO technique constrained by differing maximum velocities.

The maximum velocity that is sampled in Figure **7.8** has a minor effect on the performance of the PSO technique. The maximum velocity is of interest because a small maximum velocity will be detrimental to a global parameter search, whereas a large maximum could be detrimental to optimisation once a local minima has been found. A possible solution to these competing effects is to scale the maximum allowed velocity over the course of the simulation, this is explored in the next section.

7.2.1.6 Maximum velocity scaling

In addition to controlling the maximum velocity within a PSO run, the maximum velocity can be scaled linearly during the run in the same manner as the ω scaling in Section **7.2.1.2**. This is thought to help the simulation smoothly turn from performing a global search to optimising the found local

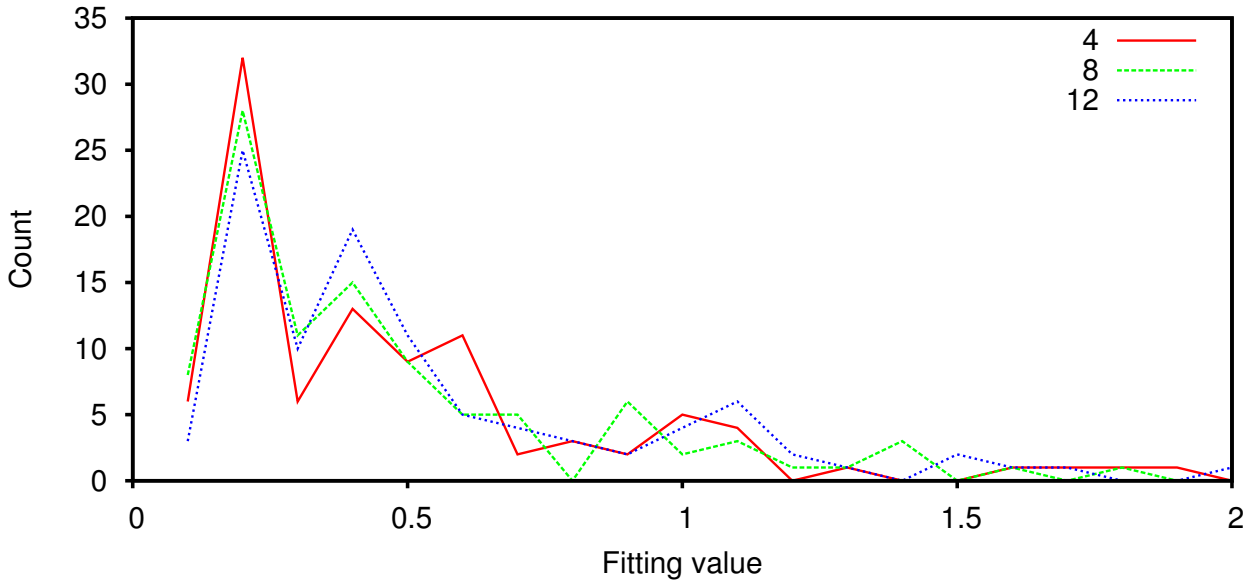


Figure 7.8: Histogram of the performance of the PSO scheme with different maximum allowed velocity, V. The legend shows the maximum allowed velocity (change in value of parameter over one iteration) for each set of simulations. (see Section **7.2.1.5**)

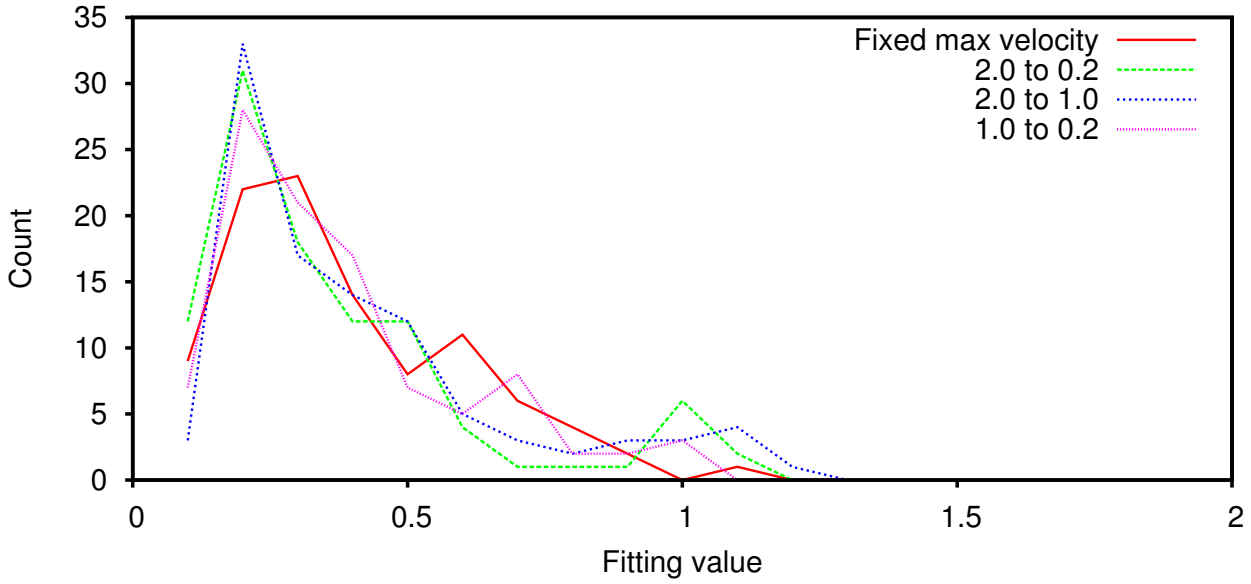


Figure 7.9: A histogram of the performance of the PSO scheme whilst using different linearly scaled maximum velocities regimes. The numbers in the legend are factors applied a maximum velocity of 8 for both parameters. When there is a range the maximum allowed velocity was linearly scaled from the larger value to the smaller value over the course of the simulation. (see Section **7.2.1.6**)

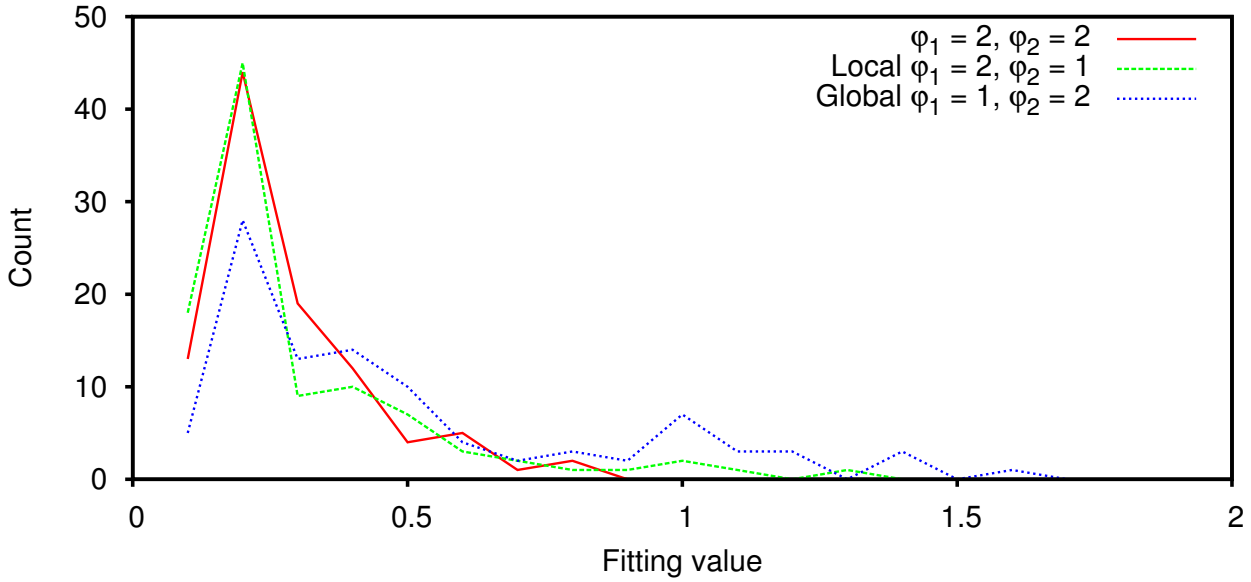


Figure 7.10: A histogram of the performance of the PSO scheme with differing values for φ_1 and φ_2. Here the global tag indicates that the swarm's history has more influence upon the velocity update and the local tag indicates the particle's history has more influence upon the velocity update. (see Section **7.2.1.7**)

minima (which may also be the global minima). Figure **7.9** shows a histogram of the performance of the PSO under a variety of velocity scaling schemes.

Figure **7.9** indicates that velocity scaling improves the efficiency of the PSO technique however it is less clear which of the scaling regimes studied offer the best improvement. It is therefore left to the specific application to find a suitable regime.

7.2.1.7 φ_1 (local) and φ_2 (global) values

Finally, the values of φ_1 and φ_2 give more options in fine tuning the PSO technique. These parameters are introduced in Eq. (**7.1.3**). φ_1 determines the influence of the local history of the particle on its velocity updates whereas φ_2 determines the influence of the swarm's history upon the update of the velocity for the particle. Figure **7.10** compares the performance of the PSO technique under different choices of the φ_1 and φ_2 parameters.

Figure **7.10** indicates that $\varphi_1 = \varphi_2$ (equal) and $\varphi_1 > \varphi_2$ (local) influence on the velocity update are largely equal in efficiency. The $\varphi_1 < \varphi_2$ (global) influence makes the PSO technique less efficient.

7.2.2 Conclusions

The key feature of the PSO technique is the swarm behaviour which emerges from the simple equations of motion detailing each particle's position and velocity update in the parameter space. However the simplicity of equations (**7.1.2**) and (**7.1.3**) leaves questions about the most efficient implementation of the PSO technique. This chapter has analysed several key features of the method to answer some of these questions, and to give some advice on the application of the PSO technique for a general optimisation problem. Section **7.2.1** describes several investigations into the sensitivity of the PSO technique to different key features of the implementation when applied to a test case as detailed in Section **7.2**. The main conclusions are:

- It is important to have multiple swarms, and for each swarm to have enough particles so that swarm behaviour emerges to take full advantage of the PSO technique.

- linear ω scaling is a key feature of the PSO technique as it improves the convergence of the method (scaling of ~ 0.9 to ~ 0.4 is suggested which is in line with previous work **[158, 159]**).

- Initial particle position has importance. It is suggested that this becomes more of an issue in a greater number of dimensions (parameters to be optimised). An estimate of the parameter being fitted can be used to speed up the optimisation process and enforce a sensible answer.

- The number of iterations is an important feature of the PSO technique. However simulations can become stuck on a non-optimal solution and so the number of iterations must be a compromise such that there is enough iterations to optimise a potential solution and the simulations being efficient such that multiple simulations can be ran.

- The maximum velocity of the particles within the PSO technique has a minor effect upon the accuracy. However introducing a linear maximum velocity scaling improves the method. There is an analogy with a reducing temperature of a real system to allow the particles to migrate to the lowest energy configuration; here the velocity is lowered allowing the particles to more effectively search a local minimum to find an optimum solution.

- The relative magnitude of φ_1 and φ_2 seem to perform equally well under the local or equal ($\varphi_1 \geq \varphi_2$) regimes (Section **7.2.1.7**). It is suggested to set $\varphi_1 = \varphi_2$ such that equal importance is

given to both the local and global aspects of the velocity update unless there is a clear problem with either the global search or convergent behaviour of the method.

This test case serves as preparation for the use of the PSO technique to optimise the parametrisation of an Empirical Valence Bond (EVB) method for performing molecular dynamics on a small cluster of sulphuric acid and water. This parametrisation work is described in the next section.

7.3 PSO application

There are two requirements for the PSO technique to be used to parameterise the EVB model, these are: data to which the model is to be fitted and secondly an implementation of the scheme.

7.3.1 Reference data

The data to which the model is fitted was obtained using DFT with the PBE [**79**] functional with a plane wave basis set, a 550 eV cut-off and a 15 Å box. The CASTEP 5.5 code [**127**] was used to determine forces on atoms in a given set of configurations (see Chapter **4** for more information on the DFT method implemented for this project). The forces data was then used as the reference case for the PSO technique. The configurations used in the fit comprised:

1. Two MD runs for trihydrated and tetrahydrated single sulphuric acid molecules in configurations described as single bead Config H and SAQH as discussed in chapters 4 and 5 (see tables **4.1** and **5.1**). The data excludes data from the pre-equilibrated period of the simulation.

2. A number of relaxed configurations designated I-n, II-n-a, III-n-a, III-n-b and III-i-a in Re *et al.* [**60**]. For each of these structures a proton was identified for transfer between a sulphuric acid and water species. This proton was repositioned on a grid of points in a cube centred at $\frac{1}{2}(\mathbf{r}_{O1} + \mathbf{r}_{O2})$, with one axis of the cube parallel to the $\mathbf{r}_{OO}$ vector, and the other two axes orthogonal but arbitrary in direction. Grid points were separated by 0.2Å giving a total of 125 points per cube. Figure **7.11** illustrates the grid of points in the I-n configuration. For each grid point an energy and force calculation was performed and the force data used for fitting the EVB model parameters.

Param.	Boundary		Velocity range		Fit
	Max	Min	Max	Min	
$V^{ij}_{const.}$	0.0	−100.0	10.0	0.2	−72.20998
P	1.0	0.01	0.1	0.002	0.50743
k	30.0	1.0	3.0	0.06	15.64862
D_{OO}	8.0	0.1	0.8	0.016	4.18888
β	8.0	0.1	0.8	0.016	2.37963
R^0_{OO}	8.0	0.1	0.8	0.008	2.46345
P'	30.0	1.0	3.0	0.06	15.41834
α	30.0	5.0	3.0	0.06	17.42592
τ_{xy}	15.0	0.5	1.5	0.03	11.55606
Δ	800.0	400.0	20.0	0.05	558.40454

Table 7.1: Table detailing the parameters which were fitted using the PSO method. The column titled 'Boundary' shows the range explored for each parameter in the PSO simulation. During each simulation the maximum allowed velocity was linearly scaled from Max to Min values under the column labelled 'Velocity range'. The values for the columns labelled 'Boundary' and 'Velocity range' were set by hand. The 'Fit' column displays the optimised values for each parameter.

7.3.2 Application

The reference data set consisted of 400 randomly picked configurations from the two sources described in the previous section. The PSO method was run 200 times with 50 particles which was chosen as a compromise of computational expense verses statistical noise. Each simulation was run for 50 steps as this was found to be the required number of steps for the parameters to have converged. For the duration of the simulation the value of ω in (7.1.1) was linearly scaled from 0.9 down to 0.4 as suggested in Section **7.2.2**. β_1 and β_2 are uniformly distributed random numbers in the range $0 \leq \beta < 1$. ϕ_1 and ϕ_2 were both given the value of 2. Each run optimises ten parameters of the model, and details of the range searched for each parameter and the optimised values are given in Table 7.1. Three parameters were not optimised. Firstly r^0_{OO} was set to zero as it is redundant. Searching over the parameter γ was found to be problematic[4]. It was therefore decided to fix γ to the value used in the MS-EVB3 model (see Table **6.1**). The parameter τ_x was fixed by hand to 0.3Å due to issues of stability as described in Section **6.3.1**. The best fit for the EVB model parameters was taken to be the averages of the parameter sets $\{p_g\}$ obtained from the 200 runs, this was seen as a fair compromise between computational expense and accuracy.

[4]The gamma parameter was found to always parametrise to the lower limit of its allowed range. It was observed that the gamma parameter also became negative when the lower limit was negative. This was considered unrealistic as the gamma parameter in Eq. (**6.2.3**) is assumed to be positive.

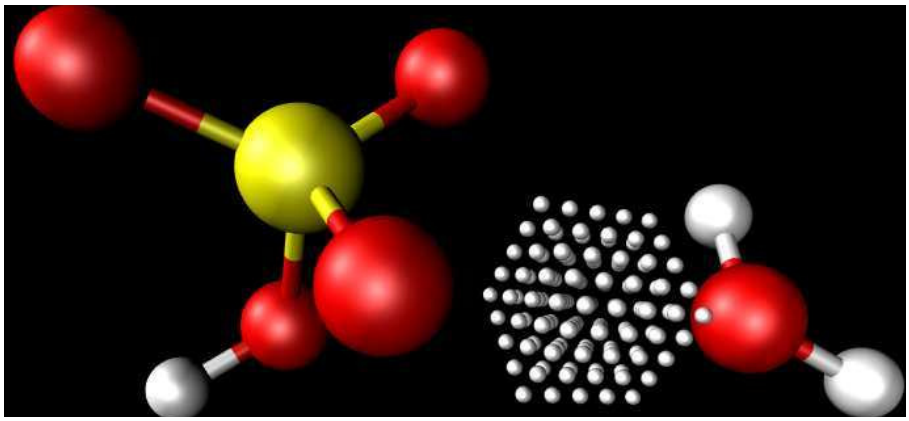
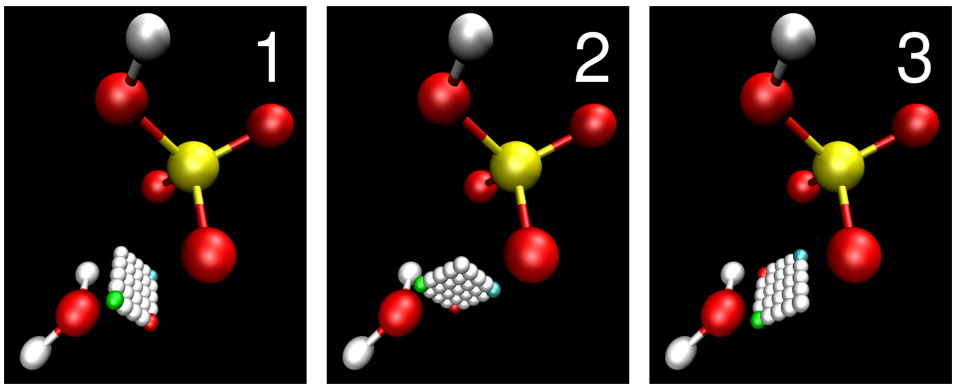


Figure 7.11: The grid of proton positions considered for configuration I-n (in the terminology of reference **[60]**) is shown. It used to generate reference force data for the PSO parameter fitting procedure.

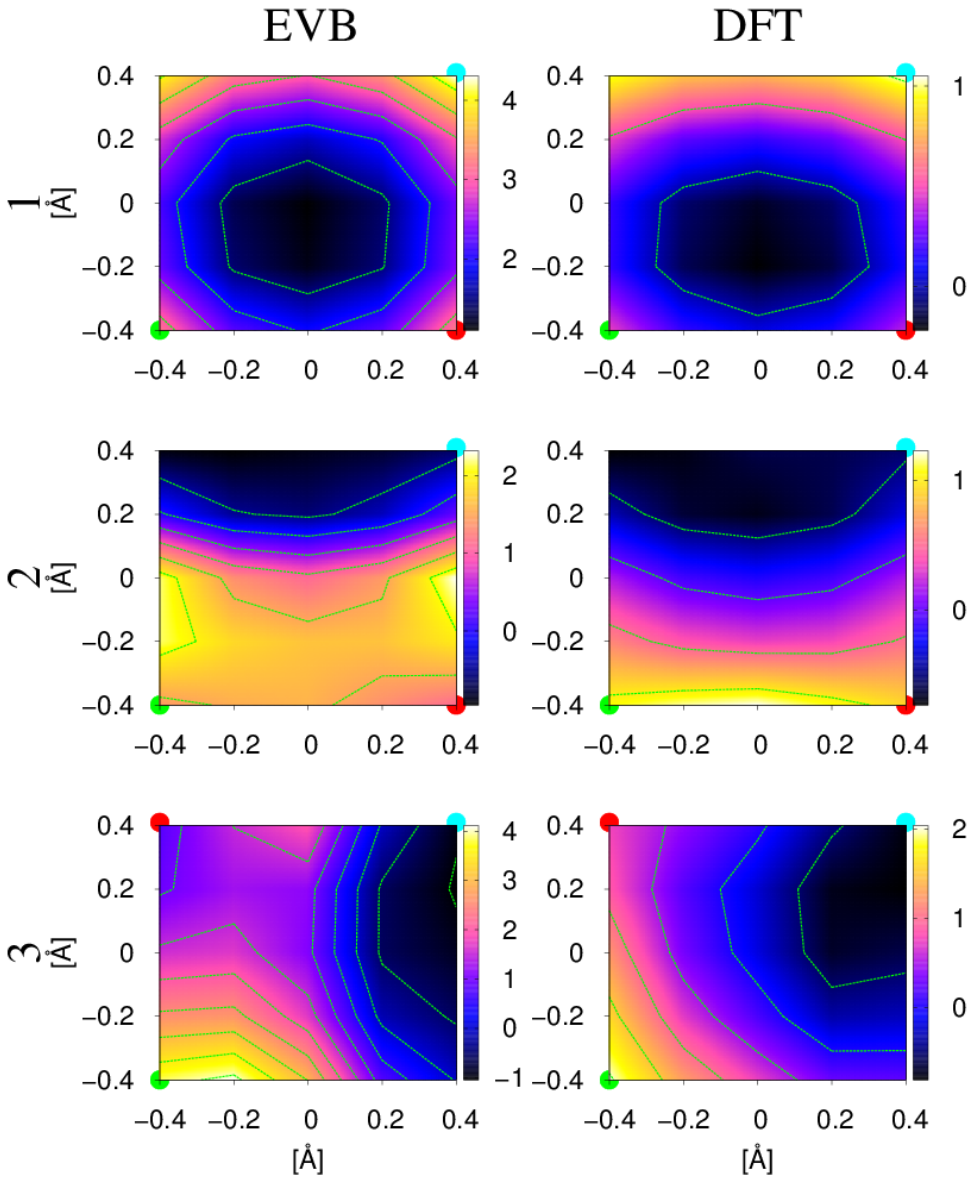
7.4 Parametrisation checks

This section focuses on assessing the accuracy of the parametrised model from the previous section in some simple cases. Figure **7.12** shows the potential energy of a sulphuric acid/water system as a function of the position of a transferring hydrogen, corresponding to the second set of reference data used in Section **7.3.1**. It is important to recognise that the EVB scheme turns off for hydrogen positions beyond a distance ± 0.3Å in the $\mathbf{r}_{OO}$ direction starting from $\frac{1}{2}(\mathbf{r}_{O1} + \mathbf{r}_{O2})$ and so energies at hydrogen positions labelled ± 0.4Å in the $\mathbf{r}_{OO}$ direction are classical and are not affected by the parametrisation scheme. The impression given is one of overall agreement between the EVB model and the DFT result, though the correspondence is not exact. The scheme is limited by the fact that it is only active for hydrogen positions lying in an ellipsoidal spatial region around the mid-point between the oxygens, that it only represents two possible bonding states per proton and that ultimately it is underpinned by a classical potential and an empirical form of the off-diagonal term. The limited basis state size reduces the ability for the system to undergo proton transfer as only under certain circumstances can it occur, an increase in the number of basis states allowed per `donor` species would be an interesting extension to this work.

The mean squared displacement, MSD is a measurement for the spatial extent of object under movement As such it is a useful tool for studying diffusion phenomena such as the diffusion of an



(a)



(b)

Figure 7.12: Contour plots of the potential energy of the I-n configuration, as labelled in reference **[60]**, as a function of the hydrogen position. The central point in the plots lies at $\frac{1}{2}(\mathbf{r}_{O1} + \mathbf{r}_{O2})$. The planes shown in (A) refer to labels 1, 2 and 3 in figure (B). The cyan, red and green spheres in (A) match the equivalent top left (cyan circle), top right (red circle) and the bottom right (green circle) of each plot in (B) respectively. The energy when the hydrogen is at position $(-0.4\text{Å}, -0.4\text{Å}, -0.4\text{Å})$ relative to the $\frac{1}{2}(\mathbf{r}_{O1} + \mathbf{r}_{O2})$ point is set to zero for both the DFT and EVB calculations. The energies are given in units of [eV].

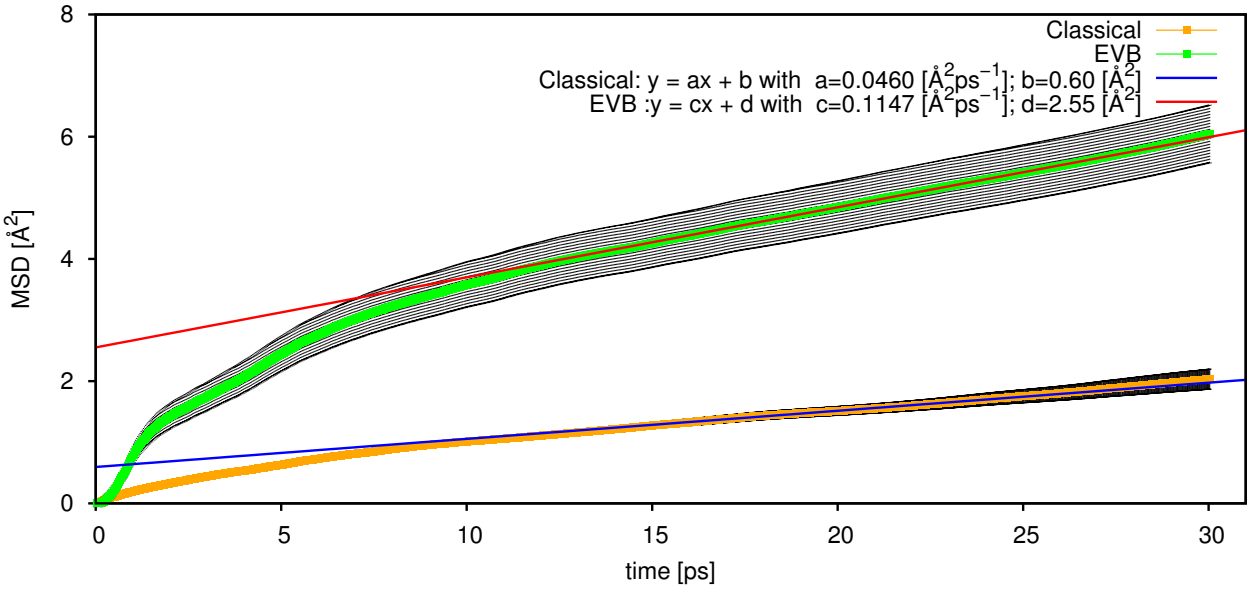


Figure 7.13: The average of the MSD value for the centre of mass for the hydronium ion over 101 simulations as a function of time. The black lines indicate the standard deviation for each value and lines are fitted to give an estimate of the diffusion coefficient. The EVB case show a clear enhancement of the diffusion of the hydronium ion in a cluster of 1 sulphuric acid and 20 waters. This is expected as the proton transfer event has a significant impact on the location of the hydronium.

excess proton in a water molecule (see reference [**103**]). The MSD is defined as,

$$\text{MSD} \equiv \left\langle (x(t) - x_0)^2 \right\rangle$$

where $x(t)$ is the position of the particle at time t and x_0 is the initial position of the particle. MSD can be used to determine the self-diffusion coefficient and hence can be used to give an indication of how the reactivity of the EVB methodology is affecting the mobility of hydroniums in the simulations presented here. Figure **7.13** shows the MSD as a function of time for a cluster of 1 sulphuric acid and 20 water molecules. The EVB run is started as a neutral sulphuric acid and quickly ionises, however as this isn't possible for the non-reactive case, this is started with the same configuration however a proton on the sulphuric acid has been attached instead to a water, and hence is started as a $[HSO_4]^- + [H_3O]^+ + 19[H_2O]$. An enhancement factor of ~ 2.5 is seen between the EVB and the classical models and the uncertainty is also enhanced. This is due in part to the inconsistency and the larger jumps in hydronium position experienced with a proton transfer. The hydronium self diffusion rate can be estimated as $1/6$th of the gradient in Figure **7.13** which gives $0.02\,[\text{Å}^2\text{s}^{-1}]$, this figure only serves to give an estimate of the diffusion of the excess proton as a more detailed

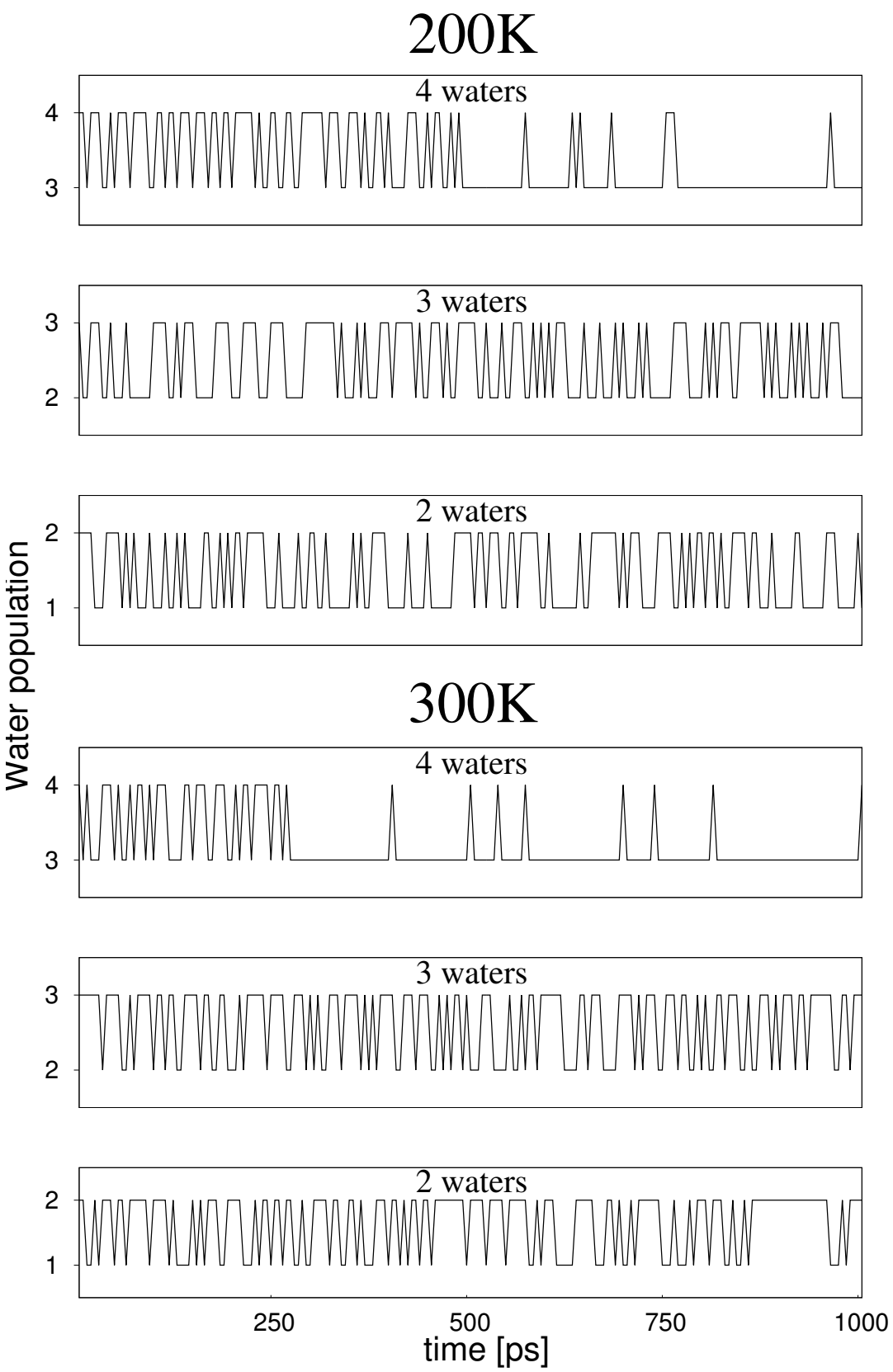


Figure 7.14: Evolution in the population of water molecules in simulations of $[H_2SO_4] + n\,[H_2O]$ where n is the number stated in the title of the plot. A decrease of one in the water population indicates an ionisation event has occurred resulting in a hydronium molecule being created (i.e. the reaction $[H_2SO_4] + [H_2O]_n \rightarrow [HSO_4]^- + [H_3O]^+ + [H_2O]_{n-1}$). The simulations were performed at 200 K and 300 K. A geometric test was used to decide the number of water molecules in the system. This causes the instabilities seen in the plot, however the large ionised periods in the plots do indicate that a stable ionisation has occurred.

analysis is needed that removes the hydronium molecule and focuses on the excess proton. However it is encouraging to see a substantial increase in mobility of the hydronium when proton transfer is introduced via the model developed here.

Figure **7.14** tracks the population of water molecules in various simulations of $[H_2SO_4] + [H_2O]_n$, where $n = 2 - 4$, over a time period of 1 ns after a 20 ps equilibration period. The simulations used a Langevin thermostat with target temperatures of 300 K or 200 K under the velocity Verlet integration scheme of a modified version of the DL_POLY 4.03 program **[118]**. There was no constraint on the centre of mass motion or the rotation of the clusters. The results indicate that at 300 K the $n = 2$ and $n = 3$ water cases see a fluctuating water population which indicate that the sulphuric acid is considering, but has not as of yet gone through with a proton transfer event. The $n = 4$ water system first shows a change in its configuration and then it becomes predominately ionised. When the same simulations are run at 200 K we see that the $n = 4$ case stays neutral for a longer period of time suggesting that the lower temperature has reduced the speed at which the system reconfigures which is as expected. These conclusions are in agreement with work performed at a higher level of theory **[58, 60, 65]** where the first ionisation event of the sulphuric acid occurs at a hydration level of between 3 and 6 water molecules. Several features are observed such as Grotthus shuffling and changes in configuration, which are expected due to the liquid-like nature of the sulphuric acid and water clusters and the complex behaviour of transferable protons within a water network. Further analysis could use the energy of the diabatic state to determine the water population in the simulation, it was not possible to complete this within the time constraints on the Ph.D. project.

Chapter 8

Free energy calculation

In this chapter a scheme for calculating free energies via thermodynamic integration is developed for studying the nucleation of a cluster from a vapour and applied to the potential developed in the previous two chapters. The approach is explored for some simple systems.

8.1 Theory

The parameter used for the integration in this model was the temperature of the system controlled through the thermostat (see Section **3.5**). The primary interest of the model is to calculate the free energy of formation of a cluster which can be alternatively defined as the free energy difference between non-interacting molecules and interacting molecules which have condensed within a confining volume. First, an analytic expression is found for the free energy of cooling a system of non-interacting particles confined within a volume from a reference state to a target state. The thermodynamic integration method is then used to obtain an numerical expression for the interacting case of the previous calculation. It is then shown that it is possible to obtain an expression to determine the difference in free energy between that of the interacting and non-interacting cases.

8.1.1 Single molecule system

We start the derivation of the thermodynamic integration by defining the ideal case with a simple system containing one molecule of N atoms. The molecule possesses a series of harmonic vibrations known as normal modes of the molecule. The system can be split into three main contributions, translations and rotations of the molecule and internal vibrations, it is then possible to construct the

following Hamiltonian for this system **[12, 120]**,

$$\mathcal{H} = \left\{ \sum_i^3 \left[\frac{\mathcal{P}_i^2}{2\mathcal{M}} \right] \right\}_{trans} + \left\{ \sum_i^3 \left[\frac{1}{2} \left(\omega_{rot}^\top I \omega_{rot} \right)_i \right] \right\}_{rot} + \left\{ \sum_j^{3N-6} \left[\frac{p_j^2}{2m_j} \right] + \sum_j^{3N-6} \left[\frac{1}{2} m_j \omega_j^2 x_j^2 \right] \right\}_{vib} \quad (8.1.1)$$

here $\mathcal{P}$ is the net momentum for the molecule, $\mathcal{P}_i = \sum_j^N p_{j,i}$ where $p_{j,i}$ describes the i^{th} component of the momentum of atom j and $\mathcal{M}$ is the mass of the molecule, $\mathcal{M} = \sum_i^N m_i$ where m_i indicates the mass of atom i. I is the moment of inertia tensor for the molecule and ω_{rot} is the angular velocity of the molecule. p_j is the momentum for the vibration j with the conjugate position x_j, mass m_j and frequency ω_j. For non-linear molecules there are $3N-6$ normal mode contributions to the Hamiltonian **[120]**. The partition function can then be written as,

$$\mathcal{Q} = q_{trans} q_{rot} q_{vib} = \int e^{-\beta H} \Pi_i^3 \left[d\mathcal{P}_i \right] \Pi_i^3 \left[d\omega_i \right] \Pi_j^{3N-6} \left[dp_j \right] \Pi_i^3 \left[d\mathcal{X}_i \right] \Pi_i^3 \left[rotation \right] \Pi_j^{3N-6} \left[dx_j \right] \quad (8.1.2)$$

here $\mathcal{X}$ is the centre of mass such that $\mathcal{X}_i = \frac{\sum_j^N m_j r_{j,i}}{\mathcal{M}}$ where i indicates the component of the centre of mass vector and $r_{j,i}$ indicates the i^{th} component of the position vector for atom j. The $1/h^{3N}$ factor has been neglected as it does not get evaluated at any point. The rotational conjugate position (represented by $\Pi_i^3 \left[rotation \right]$) for the rotational momentum has not been explicitly calculated as there is no potential applied to rotation of the molecule and hence it is a constant, which does not affect the rest of the derivation and is ignored. There is no $1/N!$ normalisation constant as the particles are classical and distinguishable. The integral identity can be used to simplify the partition function,

$$\int_{-\infty}^{\infty} \exp\left[-\alpha y^2 \right] dy = \left(\frac{\pi}{\alpha} \right)^{\frac{1}{2}}$$

such that,

$$\mathcal{Q} = \left(\frac{2\pi\mathcal{M}}{\beta} \right)^{3/2} \left(\frac{2\pi}{\beta I} \right)^{\frac{3}{2}} \left[\Pi_j^{3N-6} \left(\frac{2m_j\pi}{\beta} \right)^{\frac{1}{2}} \right] V \left[\Pi_j^{3N-6} \left(\frac{2\pi}{m_j \omega_j^2 \beta} \right)^{1/2} \right]$$

$$\mathcal{Q} = V \left(\frac{2\pi}{\beta} \right)^3 \left(\frac{\mathcal{M}}{I} \right)^{\frac{3}{2}} \left[\Pi_j^{3N-6} \left(\frac{2\pi}{\beta\omega_j} \right) \right] \quad (8.1.3)$$

where the β dependence is such,

$$\mathcal{Q} \propto \beta^{-3(N-1)}$$

We can find the average energy of this system via,

$$\langle E \rangle = -\frac{\partial \ln \mathcal{Q}}{\partial \beta} = 3(N-1)k_B T \tag{8.1.4}$$

This is in agreement with assuming that each atom contributes $\frac{3}{2}k_B T$ and each vibration contributes a further mean potential energy $\frac{1}{2}k_B T$ giving $E = \frac{3}{2}Nk_B T + \frac{1}{2}(3N-6)k_B T$.

8.1.2 M non-interacting molecular system

The next step is to expand upon the previous derivation to include M molecules. To derive the partition function we first sum the individual Hamiltonians of each (non-interacting) molecule such that,

$$\mathbb{H} = \sum_{k}^{M} \mathcal{H}_k$$

where the k is used to indicate each molecule. The partition function by extension is then (this is simplified by considering the molecules to be distinguishable),

$$\mathbb{Q} = \Pi_k^M \mathcal{Q}_k = \Pi_k^M \left[V \left(\frac{2\pi}{\beta} \right)^3 \left(\frac{\mathcal{M}_k}{I_k} \right)^{\frac{3}{2}} \left[\Pi_j^{3N-6} \left(\frac{2\pi}{\beta \omega_{k,j}} \right) \right] \right] \tag{8.1.5}$$

here N_k indicates the total number of atoms in molecule k. The temperature dependence of the partition function is then given by,

$$\mathbb{Q} \propto \beta^{-3\sum_k^M (N_k - 1)} = \beta^{-3(\mathbb{N}-M)}$$

where $\mathbb{N}$ is the total number of atoms in the system, $\mathbb{N} = \sum_k^M N_k$. The average energy can again be derived as

$$\langle E \rangle = -\frac{\partial \ln \mathbb{Q}}{\partial \beta} = 3\left(\mathbb{N} - M\right) k_B T \tag{8.1.6}$$

this expression can be thought of as each atom contributing $3k_B T$ kinetic energy and vibrational potential energy, while each defined molecule incurs a constraint on the translation and rotation of these atoms to the sum of $3k_B T$. The next step is to derive the appropriate free energy for this ideal non-interacting system.

8.1.3 Helmholtz free energy

The free energy of interest at constant temperature, number of particle and volume is the Helmholtz free energy, $\mathcal{F}$, which can be evaluated via the partition function in the following way,

$$\mathcal{F} = -k_B T \ln \mathbb{Q}$$

we first look at the dependence of $\mathbb{Q}$ on the volume, V,

$$\mathbb{Q} \propto V^M$$

now we can define the free energy difference of going from state V_1 to state V_2 at constant temperature,

$$\begin{aligned}
\Delta\mathcal{F}(V_1 \to V_2)|_V = & \int_{V_1}^{V_2} \frac{\partial \mathcal{F}(V)}{\partial V} dV \\
= & \int_{V_1}^{V_2} -k_B T \frac{\partial \ln \mathbb{Q}}{\partial V} dV = \int_{V_1}^{V_2} -k_B T \frac{1}{\mathbb{Q}} \frac{\partial \mathbb{Q}}{\partial V} dV \\
= & \int_{V_1}^{V_2} -k_B T \frac{M}{V} \frac{\mathbb{Q}}{\mathbb{Q}} dV = \int_{V_1}^{V_2} -M k_B T \frac{1}{V} dV \\
= & -M k_B T \ln \left[\frac{V_2}{V_1} \right]
\end{aligned} \tag{8.1.7}$$

This expression allows us to freely change the volume of the system. We are also interested in using a high temperature state as a reference state. The dependence of the Helmholtz free energy upon temperature is more complex, it is therefore useful to look at the quantity defined by,

$$\frac{\mathcal{F}}{T} = \mathcal{F}_T = -k_B \ln \mathbb{Q}$$

The partition function is proportional to the temperature in the following way,

$$\mathbb{Q} \propto T^{3(\mathbb{N}-M)}$$

We can then derive the difference in the same way as before (at constant volume),

$$\begin{aligned}\Delta\mathcal{F}_T(T_1 \to T_2)|_V = & \int_{T_1}^{T_2} \frac{\partial \mathcal{F}_T}{\partial T} dT \\ = & -k_B \int_{T_1}^{T_2} \frac{\partial \ln \mathrm{Q}}{\partial T} dT \\ = & -k_B 3(\mathbb{N} - M) \ln\left(\frac{T_2}{T_1}\right) \end{aligned} \tag{8.1.8}$$

The analytical expression for $\Delta\mathcal{F}_T(T_1 \to T_2)$ now needs be compared to a numerical expression for calculating this quantity for the case where the molecules are interacting with the potential described in Chapter **6** as deriving an analytical expression is non-trivial. Thermodynamic integration offers such a method for calculating this and is described in the next section.

8.1.4 Thermodynamic integration

To calculate the free energy difference from simulation we start first with the definition of the Helmholtz free energy,

$$\mathcal{F} = U - TS$$

where U is the total internal energy and S is the entropy, we then note (at constant volume),

$$\frac{\partial \mathcal{F}_T}{\partial T} = -\frac{U}{T^2} \tag{8.1.9}$$

We can now calculate the difference in Helmholtz free energy between two states where the temperature has moved from T_1 to T_2 as,

$$\Delta\mathcal{F}_T\left(T_1 \to T_2\right) = \int_{T_1}^{T_2} -\frac{U}{T^2} dT \tag{8.1.10}$$

8.1.5 From $\mathcal{F}_T$ to $\mathcal{F}$

We are interested in the free energy difference between the interacting (cluster, $\mathcal{F}^c$) case and the non-interacting case (ideal, $\mathcal{F}^i$) however so far we have found the quantity $\mathcal{F}_T$. Therefore a way must be found to relate this quantity to $\Delta\mathcal{F} = \mathcal{F}^c - \mathcal{F}^i$. For convenience we define $\Delta\mathcal{F}_T^c$ for the interacting

case,

$$\Delta\mathcal{F}_T^c = A_c$$

and $\Delta\mathcal{F}_T^i$ for the ideal case,

$$\Delta\mathcal{F}_T^i = A_i$$

To find the quantity $\Delta\mathcal{F}$ it is convenient to state define this quantity as,

$$\Delta\mathcal{F}_T = \frac{\mathcal{F}_2}{T_2} - \frac{\mathcal{F}_1}{T_1}$$

We then define $\Delta\Delta\mathcal{F}_T = \Delta\mathcal{F}_T^i - \Delta\mathcal{F}_T^c$ which can also be expressed as,

$$\Delta\Delta\mathcal{F}_T = \frac{\mathcal{F}_2^i - \mathcal{F}_2^c}{T_2} - \frac{\mathcal{F}_1^i - \mathcal{F}_1^c}{T_1}$$

However, in the initial high temperature state the free energies are equivalent[1] and therefore $\mathcal{F}_1^i - \mathcal{F}_1^c \approx 0$ leading to the following expression,

$$\Delta\Delta\mathcal{F}_T = \frac{\mathcal{F}_2^i - \mathcal{F}_2^c}{T_2} = \frac{\Delta\mathcal{F}}{T_2}$$

and finally we find,

$$\Delta\mathcal{F} = T_2\Delta\Delta\mathcal{F}_T = T_2\left(A_i - A_c\right) \tag{8.1.11}$$

and therefore we can evaluate $\Delta\mathcal{F}$ as required. Eq. (**8.1.7**) can then be used to make a comparison with an arbitrarily dense ideal non-interacting system.

8.2 Reference state

The reference state for the system was chosen to be 1500 K as this was considered to be sufficient to make the effect of the intermolecular potential negligible. To test this assumption the average system energy over the temperature squared $\langle E\rangle / T^2$ was compared between the ideal (non-interacting case given by 8.1.6) against simulation (interacting case) for a system of 1 sulphuric acid and 4 waters within a hard sphere of radius 6.5Å, each point is calculated from the average total system energy

[1]The high temperature reference state is picked for the reason that interactions are negligible in this temperature regime.

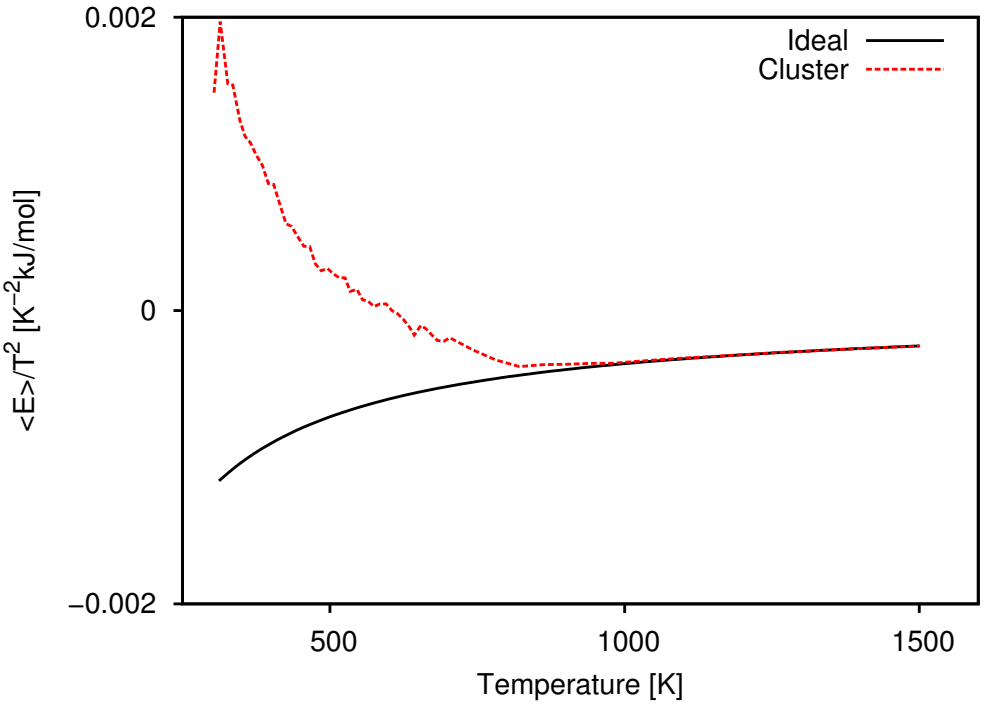


Figure 8.1: The average energy over the temperature squared is compared between the non-interacting (theory) and interacting (simulation) case. Over the temperature of 1000 K the two curves are found to be in agreement which supports the view of 1500 K being a suitable reference case.

over 1.95 *ns*, the target temperature was scaled between 1500 K and 700 K in steps of 40 K and from 700 K and 300 K in steps of 10 K. The agreement between the simulations and the theory over 1000 K shown in Figure **8.1** gives evidence that the use of 1500 K as a reference temperature is sufficient for this system.

8.3 Thermodynamic integration simulations

This section provides the preliminary results of thermodynamic integration using the EVB model derived in previous chapters.

8.3.1 Simulation details

The modified version of the DL_POLY 4.03 code was run on the Legion super computer at UCL as serial jobs[2]. Simulations used the Langevin thermostat as described in Section **3.5**. A hard enclosing sphere was implemented using a hard wall approximation where collisions with the wall reversed

[2]It is possible to develop the code to run in parallel, however this was not realised in the time scale of the project.

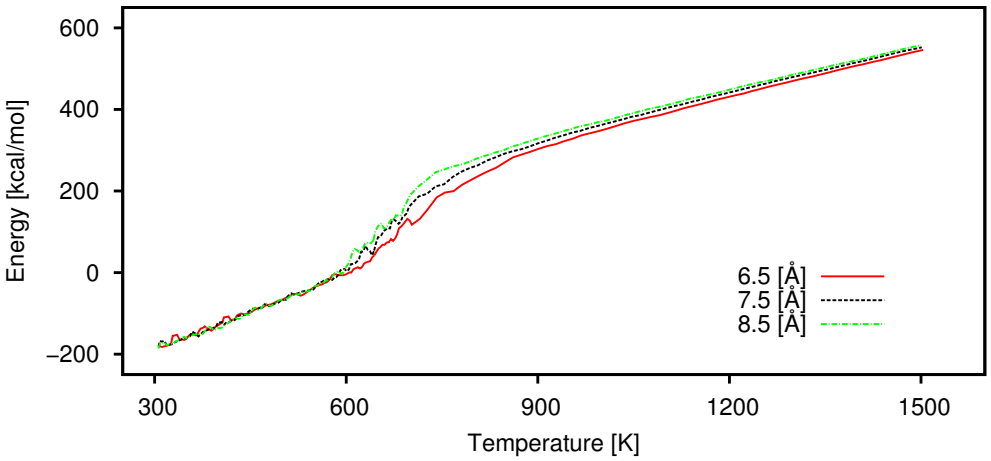


Figure 8.2: A graph of the average energy of a system as a function of the temperature. The three lines indicate differing radii of the containing hard sphere. The differing volumes affect the average energy at high temperature as this is dependent on the volume, and the middle region where the gas condenses to a cluster.

the velocity of the particle[3]. A bash script automated the submission of simulation runs, this allowed full control over the length of each step (in temperature) and the number of repetitions of each steps. Figure **8.2** shows three different thermodynamic integrations for differing hard sphere radii, each performed steps of 40 K from 1500 K to 700 K and then in steps of 10 K from 700 K to 300 K. A single run would take on the order of a couple of minutes in series, the whole script typically took 1-3 hours. The average energy for each simulation was calculated over 1.95 ns with 50 ps of equilibration, each step consisted of three simulations and a moving average over three steps was used to smooth the curve and simulations that failed[4] were removed for consistency. The middle part of the graph represent the phase transition from the liquid (cluster) phase to the gaseous phase, the difference in volume affects the free energy landscape which in turn affects the transition between gaseous and liquid phases in the system as evident from Figure **8.2**. Therefore the choice of volume can affect the difference in free energy from the high temperature state to the low temperature state, the full implications of this are not fully understood.

[3]In practice this was achieved by splitting the time-step into two parts. First allowing the particle to progress to the wall of the containing sphere, then reversing the velocity and allowing it to spend the rest of the time-step to travel in the reverse direction. In the future a scheme where the particles would be reflected at the boundary of the hard sphere could be explored.

[4]There are some minor issues with the code that caused these failures, typically the number of simulations that failed was in the order of 1% of the total ran.

8.4 Preliminary results

The goal of this study is to calculate the free energy landscape of the formation of a cluster of sulphuric acid and water. To do this a model was developed (see Chapter **6**) for use with thermodynamic integration as discussed earlier in this chapter. However due to constraints on time on this project only preliminary results were obtained with questions yet to answer. This section details the results gained thus far.

Figure **8.3** shows the results of six thermodynamic integration simulations as described in the previous section except for the simulation length, the length of these simulations was 2.95 ns, all simulations were ran with a radius of 8.5 Å. Each point on the curve was repeated five times and a running mean was used to smooth the data. It is interesting to note that above 800 K the correlation is linear and is generally smoother than that below 800K. This is due to the system becoming a gas and therefore the potential energy's contribution to the total energy becomes increasingly negligible in comparison to the kinetic energy's contribution. The kinetic energy is well controlled by the thermostat and therefore this behaviour is expected, and is vital for picking a suitable reference case where the simulation is well represented by an ideal non-interacting gas of molecules.

An interesting feature is the change in the clustering behaviour which indicates that the larger clusters are more likely to cluster at a higher temperature than the smaller clusters. This is inferred from the rightwards shift in the in the step feature observed in the graphs between the high temperature and low temperature regimes. More investigation is needed to understand this behaviour.

Figure **8.4** displays the difference in free energy from an ideal non-interacting case to that of the interacting case. This is the equivalent of the sum of the free energy cost of creating a surface for the cluster and the volumetric gain of having a bulk liquid. The figure shows that for the first four hydrations of the sulphuric acid there is a net free energy cost in producing these cluster, however this rate appears to be slowing. This shows the feature of the classical nucleation barrier as discussed in Section **1.5** and could be used to calculate the nucleation rate, however this is preliminary work and as such there are still questions to answer before it would be suitable to do so. These include the effect of volume on the simulations and the relationship between these free energies difference and the free energy difference used for CNT calculations (see Section **1.5**).

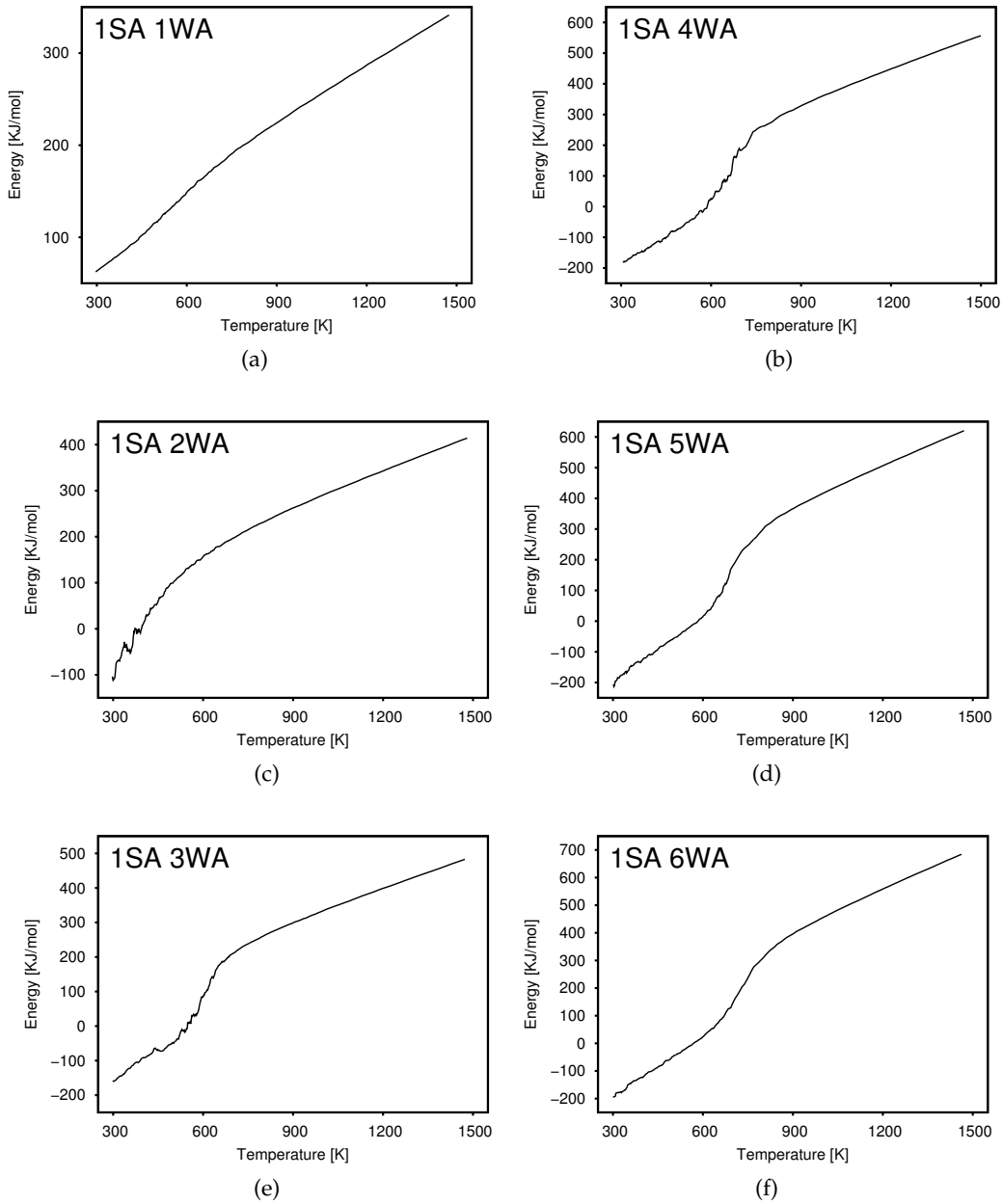


Figure 8.3: Thermodynamic integrations of system consisting of $[H_2SO_4] + m[H_2O]$ with $m = 1 - 6$. The key in the top left of each plot indicates the species involved in the thermodynamic integration using SA (sulphuric acid) and WA (water).

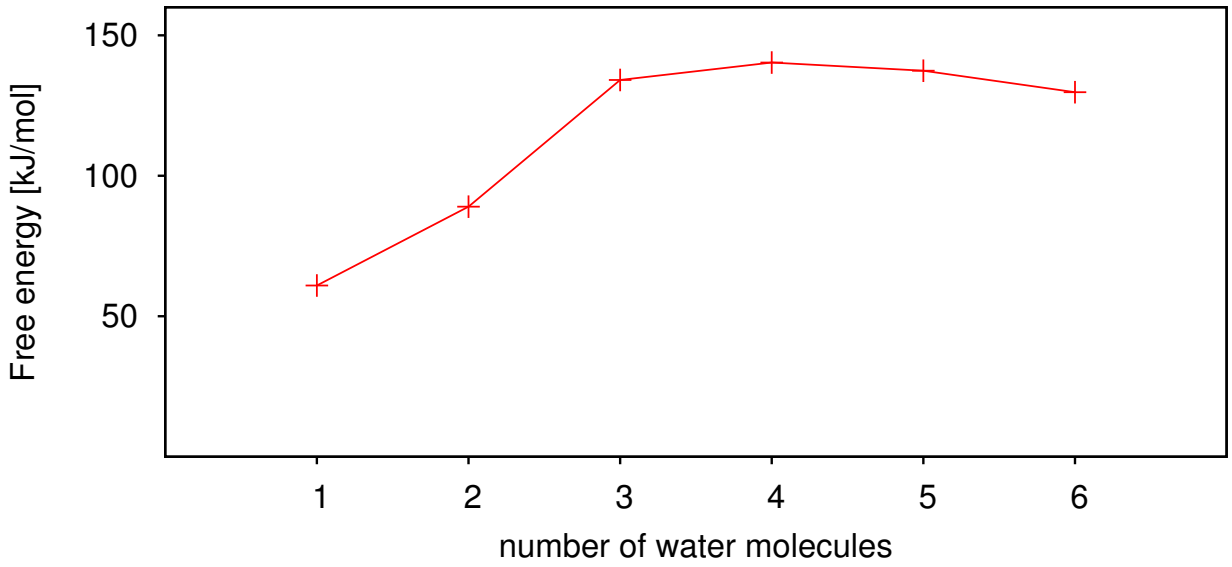


Figure 8.4: Free energy difference between ideal non-interacting case and the interacting case as calculated via the thermodynamic integration method, this is the quantity referred to as $\Delta\mathcal{F}$ in Eq. (**8.1.11**).

8.5 Conclusion

This chapter constructed an appropriate thermodynamic integration method for use with the EVB model developed in earlier chapters. Some earlier work is performed to show the method being used and some positive results are obtained (see Figures 8.3 and 8.4) however it is clear that there are still unanswered questions and work is still ongoing in this aspect of the project.

Chapter 9

Conclusion

This chapter concludes the thesis with discussion of the work performed and the implications regarding the main focus of improving the modelling of sulphuric acid and water. In addition a road map is provided for a general molecular dynamics code that takes the elements introduced by the EVB methodology to provide a simple reactive empirical code for performing molecular simulations.

9.1 Quantum chemistry

In Chapters **4** and **5** the inclusion of zero-point motion within the sulphuric acid and water system was investigated. Several configurations were studied with special attention being given to the hydrogen atoms involved in hydrogen bonds between the sulphuric acid and waters as these were suspected to be the more susceptible to zero-point motion. A trihydrated sulphuric acid configuration that was thought to be the ionised under standard geometry optimisation arguments [**58**, **60**, **69**] was investigated. The structure was found to be less clear under DFT MD as it exhibited both ionised and neutral behaviour. Further work on this structure found that there was correlation between the existence of the other two sulphuric-water hydrogen bonds and the ionisation of the system. At 300 K these bonds were found to be unstable causing the instability of the ionised state of the system. Using the harmonic approximation with this structure would likely be erroneous at 300 K due to effects observed in this study.

The addition of zero-point motion was found to have a minor effect on hydrogen bonds in a trihydrated and tetrahydrated system. In general zero-point motion was found to increase the level of ionisation in the systems in line with work performed by Li *et al.* [**146**]. Molecules such as amines and ammonia are also thought to be of significance for atmospheric nucleation and studies of the

importance of zero-point motion in these systems would be of interest.

The computational expense of these quantum chemistry simulations made it clear that it would not be feasible under current technology to evaluate cluster free energies which incorporated liquid-like behaviour using quantum chemistry methods. For this reason alternative approaches where investigated. The EVB method was chosen as it is a reactive empirical approach that was relatively simple to implement within the time scale of this project.

9.2 EVB development

The development of the EVB model has involved several key milestones, which includes,

- The creation of algorithms to identify the basis states of interest in a simulation. This required ongoing refinement to improve the performance of the algorithm.
- The implementation of the EVB method as a third party extension of the DL_POLY 4.03 source code.
- The parametrisation of the model for sulphuric acid and water proton transfer.
- Checks performed on the model to ascertain that key features were observed from simulations.

The EVB approach allows for time scales of the order of ns to be studied (for systems of the order of 100 atoms). Such access to larger time scales than is possible under current quantum chemistry techniques is helpful for studying rare events. The code required a significant amount of development and was very limited in its scope as it was written only for the sulphuric acid and water system. The next section gives a basic outline of an EVB code that would allow for a quicker implementation.

9.3 Road map for a general EVB MD code

The EVB methodology was identified in this project as a simple solution to adding the ability for reactions to occur with an underlying empirical potential. In the course of the project it became clear that this approach was very general in its reach and could easily be extended to encompass different systems, however the code written in this project is specifically for the sulphuric acid and water system. This section provides a draft plan for a code where the addition of reactions is treated in

such a way that making a system reactive is simplified to specifying the off-diagonal potential and the species involved in the reaction.

9.3.1 Components of the code

The EVB approach is underpinned by an empirical non-reactive model. This part is provided by numerous code packages (an example of which is the DL_POLY code) and is not discussed at length here, however these codes are generally made of three components,

- The configuration of the system to be simulated.
- The definition of the potential.
- Simulation details.

The first advice provided in this section is that two additional pieces of information are required, namely,

- Definition of EVB off-diagonal term.
- Definition of which species a proton can transfer between.

A more advanced treatment may require more information such as the number of basis states to be sampled and rules for how species can participate in proton transfers.

9.3.2 Structure of the code

It is suggested that the underlying architecture of the code, should take advantage of object orientated programming as to make the code significantly more readable. The three main objects may be of the form,

- Atom
- Molecule
- Basis state

This structure allows the creation of multiple basis states (definition of bonded atoms within the system) in a straight forward manner and has the advantage that the ground state selector algorithm (see Section **6.3.2.1**) is no longer required for the simulation. The molecule object simplifies extensively the application of the empirical potential and the addition calculations required for the EVB methodology. The atom object stores the basic information for each atom such as position, velocity, force and mass.

9.3.3 Application

The EVB method can be adapted for any reaction that would occur within a molecular dynamics simulation. It can be used to control which atoms can be transferred and the potential energy surface they experience during the transfer and is not limited to just sulphuric acid as been studied here. It is suggested here that a purpose built molecular dynamics code would allow a much wider adoption as it would significantly reduce the effort needed to implement the method. This would ideally be able to cope with multiple `donor` species and would be flexible in the parametrisation of the off-diagonal term (i.e. changing the form of empirical function that is used to calculate the off-diagonal term.

9.4 Preliminary conclusions

The results gained in the Chapter **8** point towards a nucleation barrier as there is an initial cost to adding waters to sulphuric acid. However there are still unanswered questions which due to the constraints on time this will be left open as of the time of writing, these questions include:

- The corrections required to ensure the right density is used for the simulations.
- Ideally addition simulations would be run to assess the sensitivity of the simulations on the initial configuration (this should only affect the lower temperature range).
- The smoothness of the curves should be addressed by additional simulations.

Once these questions have been answered satisfactorily then a nucleation rate can be acquired for a given density and temperature, the results gained in Chapter **8** show proof of principle. hat this can be used to calculate nucleation rates via the classical nucleation rate with improved free energy calculations via molecular dynamics with the thermodynamic integration technique.

9.5 Final summary

While the results from the thesis are still in an early state there are several achievements that have been made during the project. These include:

- Study of zero-point motion in small clusters of sulphuric acid and water. Results indicate that there is a small but measurable effect.
- The creation of a reactive model for sulphuric acid and water.
- Programming the code as an extension to DL_POLY. The code will be made available on github website.
- Programming a lightweight code for a PSO implementation. The code will be made available on github website.
- Development of a thermodynamic integration framework for use with the model.
- Some preliminary results that indicate that the work is making positive progress.

The unanswered questions concerning how to interpret the thermodynamic integration results are open for further research. These are left as open questions at the end of the thesis.

Bibliography

[1] J. L. Stinson, S. M. Kathmann and I. J. Ford "Dynamical consequences of a constraint on the Langevin thermostat in molecular cluster simulation" *Mol. Phys.* , pp. 1–4 (2014). doi: **10.1080/00268976.2014.917732** (see p. **1**)

[2] J. L. Stinson, S. M. Kathmann and I. J. Ford "Investigating the significance of zero-point motion in small molecular clusters of sulphuric acid and water." *J. Chem. Phys.* **140**(2), 024306, (2014). doi: **10.1063/1.4860973** (see p. **1**)

[3] S. Solomon "Stratospheric ozone depletion: A review of concepts and history" *Rev. GeoPhy.* **37**(3), pp. 275–316 (1999). doi: **10.1029/1999RG900008** (see p. **2**)

[4] J. Sun and P. Ariya "Atmospheric organic and bio-aerosols as cloud condensation nuclei (CCN): A review" *Atmos. Environ.* **40**(5), pp. 795–820 (2006). doi: **10.1016/j.atmosenv.2005.05.052** (see p. **2**)

[5] J. G. Shepherd *Geoengineering the climate: science, governance and uncertainty*. 10/29. Royal Society, (2009), url: **eprints.soton.ac.uk/156647/** (see pp. **2**, **3**)

[6] R. von Glasow, N. Bobrowski and C. Kern "The effects of volcanic eruptions on atmospheric chemistry" *Chem. Geol.* **263**(1-4), pp. 131–142 (2009). doi: **10.1016/j.chemgeo.2008.08.020** (see p. **2**)

[7] R. Zhang, A. Khalizov, L. Wang, M. Hu and W. Xu "Nucleation and growth of nanoparticles in the atmosphere" *Chem. Rev.* **112**(3), pp. 1957–2011 (2012). doi: **10.1021/cr2001756** (see pp. **3**, **14**, **20**)

[8] M. Sipilä *et al.* "The role of sulfuric acid in atmospheric nucleation" *Science (New York, N.Y.)* **327**(5970), pp. 1243–1246 (2010). doi: **10.1126/science.1180315** (see pp. **3**, **15**)

[9] R. Zhang "Atmospheric science. Getting to the critical nucleus of aerosol formation" *Science*, **328**(5984), pp. 1366–1367 (2010). doi: **10.1126/science.1189732** (see p. **3**)

[10] J. Kirkby *et al.* "Role of sulphuric acid, ammonia and galactic cosmic rays in atmospheric aerosol nucleation" *Nature*, **476**(7361), pp. 429–433 (2011). doi: **10.1038/nature10343** (see pp. **3**, **15**, **38**, **75**)

[11] O. Kupiainen, I. K. Ortega, T. Kurtén and H. Vehkamäki "Amine substitution into sulfuric acid - ammonia clusters" *Atmos. Chem. Phys.* **12**(8), pp. 3591–3599 (2012). doi: **10.5194/acp-12-3591-2012** (see p. **3**)

[**12**] M. E. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulations*. Oxford University Press, (2010), p. 720 (see pp. **4**, **22**, **30**, **31**, **33**, **36**, **60**, **61**, **63**, **66**, **114**)

[**13**] H. Vehkamäki, *Classical Nucleation Theory in Multicomponent Systems*. Springer, (2006) (see p. **6**)

[**14**] S. M. Kathmann "Understanding the chemical physics of nucleation" *Theor. Chem. Acc.* **116**(1-3), pp. 169–182 (2005). doi: **10.1007/s00214-005-0018-8** (see p. **6**)

[**15**] R. Becker and W. Döring "Kinetische Behandlung der Keimbildung in übersättigten Dämpfen" *Annalen der Physik*, **416**(8), pp. 719–752 (1935). doi: **10.1002/andp.19354160806** (see pp. **6**, **8**, **12**)

[**16**] J. Zeldovich "Theory of the formation of a new phase" *J. Expl. Theoret. Physics (USSR)*, **12**(525), (1942) (see p. **6**)

[**17**] J. I. Steinfeld, J. S. Francisco and W. L. Hase, *Chemical Kinetics and Dynamics* (2nd Edition). Prentice Hall, (1998), p. 560 (see p. **6**)

[**18**] V. I. Kalikmanov, *Nucleation Theory (Lecture Notes in Physics)*. Springer, (2012), p. 300 (see pp. **10**, **12**)

[**19**] J. E. McDonald "Homogeneous Nucleation of Vapor Condensation. I. Thermodynamic Aspects" *Am. J. Phys.* **30**(12), pp. 870–877 (1962). doi: **10.1119/1.1941841** (see p. **10**)

[**20**] James E. McDonald "Homogeneous Nucleation of Vapor Condensation. II. Kinetic Aspects" *Am. J. Phys.* **31**(1), pp. 31–41 (1963). doi: **10.1119/1.1969234** (see p. **10**)

[**21**] JB Zeldovich "On the theory of new phase formation, cavitation" *ACTA Physicochim. URSS*, **18**(1), pp. 1 –22 (1943) (see pp. **11**, **12**)

[**22**] V. I. Kalikmanov, J. Wölk and T. Kraska "Argon nucleation: bringing together theory, simulations, and experiment." *J. Chem. Phys.* **128**(12), 124506, (2008). doi: **10.1063/1.2888995** (see p. **12**)

[**23**] M. Volmer, *Kinetik der phasenbildung*. J. W. Edwards, (1939), p. 220 (see p. **12**)

[**24**] I. J. Ford "Statistical mechanics of nucleation: a review" *Proc. Inst. Mech. Eng., Part C: J. Mech. Eng. Sci.* **218**(8), pp. 883–899 (2004). doi: **10.1243/0954406041474183** (see p. **12**)

[**25**] I. Kusaka "Statistical mechanics of nucleation: Incorporating translational and rotational free energy into thermodynamics of a microdroplet" *Phys. Rev. E*, **73**(3), 031607, (2006). doi: **10.1103/PhysRevE.73.031607** (see p. **12**)

[**26**] M. Kulmala, A. Laaksonen and S. L. Girshick "The self-consistency correction to homogeneous nucleation: Extension to binary systems" *J. Aerosol Sci.* **23**(3), pp. 309–312 (1992). doi: **10.1016/0021-8502(92)90331-O** (see p. **12**)

[**27**] S. L. Girshick and C.-P. Chiu "Kinetic nucleation theory: A new expression for the rate of homogeneous nucleation from an ideal supersaturated vapor" *J. Chem. Phys.* **93**(2), pp. 1273–1277 (1990). doi: **10.1063/1.459191** (see p. **12**)

[**28**] N. P. Rao and P. H. McMurry "Effect of the Tolman Surface Tension Correction on Nucleation in Chemically Reacting Systems" *Aerosol Sci. Tech.* **13**(2), pp. 183–195 (1990). doi: **10.1080/02786829008959436** (see p. **12**)

[**29**] L. Gránásy "Diffuse interface theory for homogeneous vapor condensation" *J. Chem. Phys.* **104**(13), pp. 5188–5198 (1996). doi: **10.1063/1.471146** (see p. **13**)

[**30**] J. W. Cahn and J. E. Hilliard "Free Energy of a Nonuniform System. I. Interfacial Free Energy" *J. Chem. Phys.* **28**(2), pp. 258–267 (1958). doi: **10.1063/1.1744102** (see p. **13**)

[**31**] J. W. Cahn and J. E. Hilliard "Free Energy of a Nonuniform System. III. Nucleation in a Two-Component Incompressible Fluid" *J. Chem. Phys.* **31**(3), pp. 688–699 (1959). doi: **10.1063/1.1730447** (see p. **13**)

[**32**] D. Reguera and H. Reiss "Fusion of the Extended Modified Liquid Drop Model for Nucleation and Dynamical Nucleation Theory" *Phys. Rev. Lett.* **93**(16), 165701, (2004). doi: **10.1103/PhysRevLett.93.165701** (see p. **13**)

[**33**] D. Reguera and H. Reiss "Extended Modified Liquid Drop-Dynamical Nucleation Theory (EMLD-DNT) Approach to Nucleation: A New Theory" *J. Phys. Chem. B*, **108**(51), pp. 19831–19842 (2004). doi: **10.1021/jp047168h** (see p. **13**)

[**34**] X. C. Zeng and D. W. Oxtoby "Gas-liquid nucleation in Lennard-Jones fluids" *J. Chem. Phys.* **94**(6), pp. 4472–4478 (1991). doi: **10.1063/1.460603** (see p. **13**)

[**35**] I. Napari, J. Julin and H. Vehkamäki "Performance of some nucleation theories with a nonsharp droplet-vapor interface" *J. Chem. Phys.* **133**(15), 154503, (2010). doi: **10.1063/1.3502643** (see p. **13**)

[**36**] J. Kazil and E. R. Lovejoy "A semi-analytical method for calculating rates of new sulfate aerosol formation from the gas phase" *Atmos. Chem. Phys. Discuss*, **7**(1), pp. 2169–2196 (2007). doi: **10.5194/acpd-7-2169-2007** (see p. **13**)

[**37**] V. I. Kalikmanov "Mean-field kinetic nucleation theory." *J. Chem. Phys.* **124**(12), 124505, (2006). doi: **10.1063/1.2178812** (see p. **13**)

[**38**] G. J. Doyle "Self-Nucleation in the Sulfuric Acid-Water System" *J. Chem. Phys.* **35**(3), pp. 795–799 (1961). doi: **10.1063/1.1701218** (see p. **14**)

[**39**] H. Vehkamäki "An improved parameterization for sulfuric acid-water nucleation rates for tropospheric and stratospheric conditions" *J. Geophys. Res.-Atmos.* **107**(D22), AAC 3–1–AAC 3–10 (2002). doi: **10.1029/2002JD002184** (see p. **14**)

[**40**] S. M. Kreidenweis, R. C. Flagan, J. H. Seinfeld and K. Okuyama "Binary nucleation of methanesulfonic acid and water" *J. Aerosol Sci.* **20**(5), pp. 585–607 (1989). doi: **10.1016/0021-8502(89)90105-5** (see p. **14**)

[**41**] S. M. Ball, D. R. Hanson, F. L. Eisele and P. H. McMurry "Laboratory studies of particle nucleation: Initial results for H_2SO_4, H_2O, and NH_3 vapors" *J. Geophys. Res.* **104**(D19), pp. 23709–23718 (1999). doi: **10.1029/1999JD900411** (see p. **14**)

[**42**] F. L. Eisele and D. R. Hanson "First Measurement of Prenucleation Molecular Clusters" *J. Phys. Chem. A*, **104**(4), pp. 830–836 (2000). doi: **10.1021/jp9930651** (see p. **15**)

[**43**] M. Kulmala, A. Laaksonen and L. Pirjola "Parameterizations for sulfuric acid/water nucleation rates" *J. Geophys. Res.* **103**(D7), pp. 8301–8307 (1998). doi: **10.1029/97JD03718** (see p. **15**)

[**44**] G. Wilemski "Composition of the critical nucleus in multicomponent vapor nucleation" *J. Chem. Phys.* **80**(3), pp. 1370–1372 (1984). doi: **10.1063/1.446822** (see p. **15**)

[**45**] T. Kurtén, V. Loukonen, H. Vehkamäki and M. Kulmala "Amines are likely to enhance neutral and ion-induced sulfuric acid-water nucleation in the atmosphere more effectively than ammonia" *Atmos. Chem. Phys. Discuss*, **8**(14), pp. 7455–7476 (2008). doi: **10.5194/acpd-8-7455-2008** (see pp. **15**, **75**)

[**46**] J. Almeida *et al.* "Molecular understanding of sulphuric acid-amine particle nucleation in the atmosphere." *Nature*, **502**(7471), pp. 359–363 (2013). doi: **10.1038/nature12663** (see pp. **15**, **17**)

[**47**] R. M. Martin, *Electronic structure: Basic Theory and practical methods*. Cambridge University Press, (2004) (see pp. **16**, **42–48**)

[**48**] Chr. Møller and M. S. Plesset "Note on an Approximation Treatment for Many-Electron Systems" *Phys. Rev.* **46**(7), pp. 618–622 (1934). doi: **10.1103/PhysRev.46.618** (see pp. **16**, **44**)

[**49**] R. J. Bartlett and M. Musiał "Coupled-cluster theory in quantum chemistry" *Rev. Mod. Phys.* **79**, pp. 291–352 (1 2007). doi: **10.1103/RevModPhys.79.291** (see p. **16**)

[**50**] E. H. Lieb and B. Simon "The Thomas-Fermi theory of atoms, molecules and solids" *Adv. Math.* **23**(1), pp. 22–116 (1977). doi: **10.1016/0001-8708(77)90108-6** (see p. **16**)

[**51**] P. Hohenberg and W. Kohn "Inhomogeneous Electron Gas" *Phys. Rev.* **136**(3B), B864–B871 (1964). doi: **10.1103/PhysRev.136.B864** (see pp. **16**, **42**, **150**)

[**52**] W. Kohn and L. J. Sham "Self-Consistent Equations Including Exchange and Correlation Effects" *Phys. Rev.* **140**(4A), A1133–A1138 (1965). doi: **10.1103/PhysRev.140.A1133** (see pp. **16**, **43**, **45**)

[**53**] J. P. Perdew and A. Ruzsinszky "Fourteen easy lessons in density functional theory" *Int. J. Quantum Chem.* **110**(15), pp. 2801–2807 (2010). doi: **10.1002/qua.22829** (see pp. **16**, **42**, **44**, **45**)

[**54**] H. Arstila, K. Laasonen and A. Laaksonen "Ab initio study of gas phase sulphuric acid hydrates containing 1 to 3 water molecules" *J. Phys. Chem. A*, **108**(3), pp. 1031–1039 (1998). doi: **10.1063/1.475496** (see pp. **17**, **18**, **22**)

[**55**] A. D. Becke "Density-functional exchange-energy approximation with correct asymptotic behavior" *Phys. Rev. A*, **38**(6), pp. 3098–3100 (1988). doi: **10.1103/PhysRevA.38.3098** (see p. **17**)

[**56**] C. Lee, W. Yang and R. G. Parr "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density" *Phys. Rev. B*, **37**(2), pp. 785–789 (1988). doi: **10.1103/PhysRevB.37.785** (see p. **17**)

[**57**] A. Jaecker-Voirol and P. Mirabel "Nucleation rate in a binary mixture of sulfuric acid and water vapor" *J. Phys. Chem.* **92**(12), pp. 3518–3521 (1988). doi: **10.1021/j100323a039** (see p. **17**)

[**58**] A. R. Bandy and J. C. Ianni "Study of the Hydrates of H_2SO_4 Using Density Functional Theory" *J. Phys. Chem. A*, **102**(32), pp. 6533–6539 (1998). doi: **10.1021/jp980270s** (see pp. **17**, **18**, **22**, **54**, **67**, **72**, **112**, **124**)

[**59**] J. C. Ianni and A. R. Bandy "A theoretical study of the hydrates of $(H_2SO_4)_2$ and its implications for the formation of new atmospheric particles" *J. Mol. Struc.-THEOCHEM*, **497**(1-3), pp. 19–37 (1999). doi: **10.1016/S0166-1280(99)00182-7** (see pp. **17**, **18**, **54**)

[**60**] S. Re, Y. Osamura and K. Morokuma "Coexistence of Neutral and Ion-Pair Clusters of Hydrated Sulfuric Acid $H_2SO_4 \cdot (H_2O)_n$ (n = 1-5)A Molecular Orbital Study" *J. Phys. Chem. A*, **103**(18), pp. 3535–3547 (1999). doi: **10.1021/jp984759x** (see pp. **17**, **18**, **22**, **54**, **55**, **67**, **72**, **106**, **108**, **109**, **112**, **124**)

[**61**] A. D. Becke "A new mixing of Hartree-Fock and local density-functional theories" *J. Chem. Phys.* **98**(2), pp. 1372–1377 (1993). doi: **10.1063/1.464304** (see pp. **17**, **45**)

[**62**] T. Kurtén, I. K. Ortega and H. Vehkamäki "The sign preference in sulfuric acid nucleation" *J. Mol. Struc.-THEOCHEM*, **901**(1-3), pp. 169–173 (2009). doi: **10.1016/j.theochem.2009.01.024** (see p. **19**)

[**63**] A. B. Nadykto, F. Yu and J. Herb "Towards understanding the sign preference in binary atmospheric nucleation" *Phys. Chem. Chem. Phys.* **10**(47), pp. 7073–7078 (2008). doi: **10.1039/b807415a** (see p. **19**)

[64] C.-G. Ding and K. Laasonen "Partially and fully deprotonated sulfuric acid in $H_2SO_4(H_2O)_n$ (n=6-9) clusters" *Chem. Phys. Lett.* **390**(4-6), pp. 307–313 (2004). doi: **10.1016/j.cplett.2004.02.112** (see p. **19**)

[65] C. Arrouvel, V. Viossat and C. Minot "Theoretical study of hydrated sulfuric acid: clusters and periodic modeling" *J. Mol. Struc.-THEOCHEM*, **718**(1-3), pp. 71–76 (2005). doi: **10.1016/j.theochem.2004.12.032** (see pp. **19**, **20**, **112**)

[66] L. J. Larson, M. Kuno and F.-M. Tao "Hydrolysis of sulfur trioxide to form sulfuric acid in small water clusters" *J. Chem. Phys.* **112**(20), pp. 8830–8838 (2000). doi: **10.1063/1.481532** (see p. **19**)

[67] A. A. Natsheh, A. B. Nadykto, K. V. Mikkelsen, F. Yu and J. Ruuskanen "Sulfuric Acid and Sulfuric Acid Hydrates in the Gas Phase: A DFT Investigation" *J. Phys. Chem. A*, **108**(41), pp. 8914–8929 (2004). doi: **10.1021/jp048858o** (see p. **19**)

[68] C.-G. Ding, K. Laasonen and A. Laaksonen "Two Sulfuric Acids in Small Water Clusters" *J. Phys. Chem. A*, **107**(41), pp. 8648–8658 (2003). doi: **10.1021/jp022575j** (see pp. **19**, **20**)

[69] B. Temelso *et al.* "Quantum mechanical study of sulfuric acid hydration: atmospheric implications" *J. Phys. Chem. A*, **116**(9), pp. 2209–2224 (2012). doi: **10.1021/jp2119026** (see pp. **19–21**, **52**, **54**, **72**, **124**)

[70] D. E. Husar, B. Temelso, A. L. Ashworth and G. C. Shields "Hydration of the bisulfate ion: atmospheric implications." *J. Phys. Chem. A*, **116**(21), pp. 5151–5163 (2012). doi: **10.1021/jp300717j** (see pp. **19**, **20**)

[71] B. Temelso, T. N. Phan and G. C. Shields "Computational study of the hydration of sulfuric acid dimers: implications for acid dissociation and aerosol formation." *J. Phys. Chem. A*, **116**(39), pp. 9745–9758 (2012). doi: **10.1021/jp3054394** (see pp. **19**, **20**)

[72] J. Herb, Y. Xu, F. Yu and A. B. Nadykto "Large hydrogen-bonded pre-nucleation $(HSO_4^-)(H_2SO_4)_m(H_2O)_k$ and $(H_SO_4-)(NH_3)(H_2SO_4)_m(H_2O)_k$ clusters in the earth's atmosphere." *J. Phys. Chem. A*, **117**(1), pp. 133–152 (2013). doi: **10.1021/jp3088435** (see pp. **19**, **20**)

[73] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh and C. Fiolhais "Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation" *Phys. Rev. B*, **46**(11), pp. 6671–6687 (1992). doi: **10.1103/PhysRevB.46.6671** (see p. **19**)

[74] N. A. Cordero and J. A. Alonso "The interaction of sulfuric acid with graphene and formation of adsorbed crystals" *Nanotechnology*, **18**(48), p. 485705 (2007). doi: **10.1088/0957-4484/18/48/485705** (see p. **19**)

[**75**] F. Weigend "RI-MP2: optimized auxiliary basis sets and demonstration of efficiency" *Chem. Phys. Lett.* **294**(1-3), pp. 143–152 (1998). doi: **10.1016/S0009-2614(98)00862-8** (see p. **19**)

[**76**] F. Haase and R. Ahlrichs "Semidirect MP2 gradient evaluation on workstation computers: The MPGRAD program" *J. Comput. Chem.* **14**(8), pp. 907–912 (1993). doi: **10.1002/jcc.540140805** (see p. **19**)

[**77**] R. Car and M. Parrinello "Unified Approach for Molecular Dynamics and Density-Functional Theory" *Phys. Rev. Lett.* **55**(22), pp. 2471–2474 (1985). doi: **10.1103/PhysRevLett.55.2471** (see p. **21**)

[**78**] Y.-K. Choe, E. Tsuchida and T. Ikeshoji "First-principles molecular dynamics study on aqueous sulfuric acid solutions" *J. Chem. Phys.* **126**(15), 154510, (2007). doi: **10.1063/1.2718526** (see pp. **21**, **22**)

[**79**] J. Perdew, K. Burke and M. Ernzerhof "Generalized Gradient Approximation Made Simple" *Phys. Rev. Lett.* **77**(18), pp. 3865–3868 (1996). doi: **10.1103/PhysRevLett.77.3865** (see pp. **21**, **45**, **50**, **106**)

[**80**] S. Cukierman "Et tu, Grotthuss! and other unfinished stories" *Biochimica et biophysica acta*, **1757**(8), pp. 876–885 (2006). doi: **10.1016/j.bbabio.2005.12.001** (see pp. **21**, **45**, **58**)

[**81**] K. E. Anderson, J. I. Siepmann, P. H. McMurry and J. VandeVondele "Importance of the number of acid molecules and the strength of the base for double-ion formation in $(H_2SO_4)_m \cdot base \cdot (H_2O)_6$ clusters" *J. Am. Chem. Soc.* **130**(43), pp. 14144–14147 (2008). doi: **10.1021/ja8019774** (see p. **21**)

[**82**] A. D. Hammerich, V. Buch and F. Mohamed "Ab initio simulations of sulfuric acid solutions" *Chem. Phys. Lett.* **460**(4-6), pp. 423–431 (2008). doi: **10.1016/j.cplett.2008.06.053** (see p. **22**)

[**83**] A. D. Boese, N. L. Doltsinis, N. C. Handy and M. Sprik "New generalized gradient approximation functionals" *J. Chem. Phys.* **112**(4), pp. 1670–1678 (2000). doi: **10.1063/1.480732** (see p. **22**)

[**84**] A. P. Bartók, M. C. Payne and G. Csányi "Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons" *Phys. Rev. Lett.* **104**(13), 136403, (2010). doi: **10.1103/PhysRevLett.104.136403** (see pp. **23**, **25**)

[**85**] A. C. T. van Duin, S. Dasgupta, F. Lorant and W. A. Goddard "ReaxFF: A Reactive Force Field for Hydrocarbons" *J. Phys. Chem. A*, **105**(41), pp. 9396–9409 (2001). doi: **10.1021/jp004368u** (see pp. **23**, **25**)

[**86**] H. Lin and D. G. Truhlar "QM/MM: what have we learned, where are we, and where do we go from here?" *Theor. Chem. Acc.* **117**(2), pp. 185–199 (2006). doi: **10.1007/s00214-006-0143-z** (see pp. **23**, **25**)

[**87**] S. C. L. Kamerlin and A. Warshel "The empirical valence bond model: theory and applications" *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**(1), pp. 30–45 (2011). doi: **10.1002/wcms.10** (see pp. **23**, **25**, **76**, **79**)

[**88**] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (Emended Edition). Dover Publications, (2010), p. 384 (see pp. **22**, **60**)

[**89**] A. Kakizaki, H. Motegi, T. Yoshikawa, T. Takayanagi, M. Shiga and M. Tachikawa "Path-integral molecular dynamics simulations of small hydrated sulfuric acid clusters $H_2SO_4 \cdot (H2O)_n$ (n=1-6) on semiempirical PM6 potential surfaces" *J. Mol. Struc.-THEOCHEM*, **901**(1-3), pp. 1–8 (2009). doi: **10.1016/j.theochem.2009.01.022** (see p. **22**)

[**90**] J. J. P. Stewart "Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements" *J. Mol. Model.* **13**(12), pp. 1173–213 (2007). doi: **10.1007/s00894-007-0233-4** (see p. **22**)

[**91**] J. A. Pople, *Approximate Molecular Orbital Theory (Advanced Chemistry)*. Mcgraw-Hill (Tx), (1970), p. 224 (see p. **22**)

[**92**] S. Sugawara, T. Yoshikawa, T. Takayanagi, M. Shiga and M. Tachikawa "Quantum proton transfer in hydrated sulfuric acid clusters: a perspective from semiempirical path integral simulations" *J. Phys. Chem. A*, **115**(42), pp. 11486–11494 (2011). doi: **10.1021/jp202380h** (see p. **22**)

[**93**] C. M. Maupin, N. Castillo, S. Taraphder, C. Tu, R. McKenna, D. N. Silverman and G. A. Voth "Chemical rescue of enzymes: proton transfer in mutants of human carbonic anhydrase II." *J. Am. Chem. Soc.* **133**(16), pp. 6223–6234 (2011). doi: **10.1021/ja1097594** (see pp. **24**, **26**)

[**94**] I. Kusaka, Z.-G. Wang and J. H. Seinfeld "Binary nucleation of sulfuric acid-water: Monte Carlo simulation" *J. Chem. Phys.* **108**(16), pp. 6829–6848 (1998). doi: **10.1063/1.476097** (see p. **23**)

[**95**] S. M. Kathmann and B. N. Hale "Monte Carlo Simulations of Small Sulfuric Acid-Water Clusters" *J. Phys. Chem. B*, **105**(47), pp. 11719–11728 (2001). doi: **10.1021/jp0116499** (see p. **23**)

[**96**] C.-G. Ding, T. Taskila, K. Laasonen and A. Laaksonen "Reliable potential for small sulfuric acid-water clusters" *Chem. Phys.* **287**(1-2), pp. 7–19 (2003). doi: **10.1016/S0301-0104(02)00971-0** (see pp. **23**, **87**)

[**97**] M. Toivola, I. Napari and H. Vehkamäki "Structure of water-sulfuric acid clusters from molecular dynamics simulations" *Boreal Environment Research*, **14**(4), pp. 654–661 (2009) (see p. **24**)

[**98**] A. Warshel and M. Levitt "Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme" *J. Mol. Biol.* **103**(2), pp. 227–249 (1976) (see p. **25**)

[**99**] A. Warshel and R. M. Weiss "An empirical valence bond approach for comparing reactions in solutions and in enzymes" *J. Am. Chem. Soc.* **102**(20), pp. 6218–6226 (1980). doi: **10.1021/ja00540a008** (see pp. **25**, **76**)

[**100**] R. Vuilleumier and D. Borgis "Quantum Dynamics of an Excess Proton in Water Using an Extended Empirical Valence-Bond Hamiltonian" *J. Phys. Chem. B*, **102**(22), pp. 4261–4264 (1998). doi: **10.1021/jp9807423** (see p. **25**)

[**101**] U. W. Schmitt and G. A. Voth "Multistate Empirical Valence Bond Model for Proton Transport in Water" *J. Phys. Chem. B*, **102**(29), pp. 5547–5551 (1998). doi: **10.1021/jp9818131** (see pp. **25**, **76**)

[**102**] T. J. F. Day, A. V. Soudackov, M. Čuma, U. W. Schmitt and G. A. Voth "A second generation multistate empirical valence bond model for proton transport in aqueous systems" *J. Chem. Phys.* **117**(12), pp. 5839–5849 (2002). doi: **10.1063/1.1497157** (see p. **25**)

[**103**] Y. Wu, H. Chen, F. Wang, F. Paesani and G. A. Voth "An improved multistate empirical valence bond model for aqueous proton solvation and transport" *J. Phys. Chem. B*, **112**(2), pp. 467–482 (2008). doi: **10.1021/jp076658h** (see pp. **25**, **80**, **81**, **86**, **93**, **110**)

[**104**] M. Čuma, Udo W. Schmitt and G. A. Voth "A Multi-State Empirical Valence Bond Model for Weak Acid Dissociation in Aqueous Solution" *J. Phys. Chem. A*, **105**(12), pp. 2814–2823 (2001). doi: **10.1021/jp0038207** (see p. **25**)

[**105**] C. M. Maupin, K. F. Wong, A. V. Soudackov, S. Kim and G. A. Voth "A multistate empirical valence bond description of protonatable amino acids." *J. Phys. Chem. A*, **110**(2), pp. 631–639 (2006). doi: **10.1021/jp053596r** (see p. **25**)

[**106**] K. Park, W. Lin and F. Paesani "A refined MS-EVB model for proton transport in aqueous environments." *J. Phys. Chem. B*, **116**(1), pp. 343–352 (2012). doi: **10.1021/jp208946p** (see p. **26**)

[**107**] G. Brancato and M. E. Tuckerman "A polarizable multistate empirical valence bond model for proton transport in aqueous solution." *J. Chem. Phys.* **122**(22), 224507, (2005). doi: **10.1063/1.1902924** (see p. **26**)

[**108**] D. Frenkel and B. Smit, *Understanding Molecular Simulations: from Algorithms to Application*. Second. Elsevier, (2001), p. 664 (see pp. **27**, **39**, **69**, **70**, **145**)

[**109**] J. M. Haile, *Molecular Dynamics Simulation: Elementary Methods (Wiley Professional)*. Wiley-Interscience, (1997), p. 512 (see pp. **27**, **28**, **30**, **54**, **143**, **145**, **148**)

[**110**] D. C. Rapaport, *The Art of Molecular Dynamics Simulation*. Cambridge University Press, (2004), p. 564 (see pp. **27**, **39**)

[**111**] N. Metropolis and S. Ulam "The Monte Carlo Method" *J. Am. Stat. Assoc.* **44**(247), pp. 335–341 (1949). doi: **10.1080/01621459.1949.10483310** (see p. **29**)

[**112**] M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods*. Wiley VCH, (2008), p. 215 (see p. **29**)

[**113**] L. Verlet "Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules" *Phys. Rev.* **159**(1), pp. 98–103 (1967). doi: **10.1103/PhysRev.159.98** (see p. **30**)

[**114**] W. C. Swope "A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters" *J. Chem. Phys.* **76**(1), pp. 637–649 (1982). doi: **10.1063/1.442716** (see p. **30**)

[**115**] S. Nosé "A unified formulation of the constant temperature molecular dynamics methods" *J. Chem. Phys.* **81**(1), pp. 511–519 (1984). doi: **10.1063/1.447334** (see p. **30**)

[**116**] W. Hoover "Canonical dynamics: Equilibrium phase-space distributions" *Phys. Rev. A*, **31**(3), pp. 1695–1697 (1985). doi: **10.1103/PhysRevA.31.1695** (see p. **30**)

[**117**] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak "Molecular dynamics with coupling to an external bath" *J. Chem. Phys.* **81**(8), pp. 3684–3690 (1984). doi: **10.1063/1.448118** (see p. **30**)

[**118**] I. T. Todorov, W. Smith, K. Trachenko and M. T. Dove "DL_POLY_3: new dimensions in molecular dynamics simulations via massive parallelism" *J. Mater. Chem.* **16**(20), pp. 1911–1918 (2006). doi: **10.1039/b517931a** (see pp. **31**, **112**)

[**119**] T. Çağin and J. R. Ray "Isothermal molecular-dynamics ensembles" *Phys. Rev. A*, **37**, pp. 4510–4513 (11 1988). doi: **10.1103/PhysRevA.37.4510** (see p. **36**)

[**120**] I. J. Ford, *Statistical Physics: an entropic approach*. Wiley, (2013), p. 280 (see pp. **37**, **114**)

[**121**] H. Flyvbjerg and H. G. Petersen "Error estimates on averages of correlated data" *J. Chem. Phys.* **91**, pp. 461–466 (1989). doi: **10.1063/1.457480** (see pp. **39**, **69**, **70**)

[**122**] R. Kumar, J. R. Schmidt and J. L. Skinner "Hydrogen bonding definitions and dynamics in liquid water" *J. Chem. Phys.* **126**, 204107, (2007). doi: **10.1063/1.2742385** (see pp. **41**, **68**)

[**123**] R. L. Hayes, S. J. Paddison and M. E. Tuckerman "Proton transport in triflic acid pentahydrate studied via ab initio path integral molecular dynamics" *J. Phys. Chem. A*, **115**, pp. 6112–6124 (2011). doi: **10.1021/jp110953a** (see p. **41**)

[**124**] J. P. Perdew, A. Ruzsinszky, L. A. Constantin, J. Sun and G. I. Csonka "Some Fundamental Issues in Ground-State Density Functional Theory: A Guide for the Perplexed" *J. Chem. Theory Comput.* **5**(4), pp. 902–908 (2009). doi: **10.1021/ct800531s** (see pp. **42**, **44**)

[**125**] M. C. Payne, T. A. Arias and J. D. Joannopoulos "Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients" *Rev. Mod. Phys.* **64**(4), pp. 1045–1097 (1992). doi: **10.1103/RevModPhys.64.1045** (see pp. **42**, **49**, **50**)

[**126**] W. Kohn "Nobel Lecture: Electronic structure of matter-wave functions and density functionals" *Rev. Mod. Phys.* **71**(5), pp. 1253–1266 (1999). doi: **10.1103/RevModPhys.71.1253** (see p. **43**)

[**127**] S. J. Clark, M. D. Segall and C. J. Pickard "First principles methods using CASTEP" *Z. Kristallogr.* **220**(5/6/2005), pp. 567–570 (2005). doi: **10.1524/zkri.220.5.567.65075** (see pp. **44**, **50**, **66**, **106**)

[**128**] J. Čížek "On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods" *J. Chem. Phys.* **45**(11), pp. 4256–4266 (1966). doi: **10.1063/1.1727484** (see p. **44**)

[**129**] A. D. Boese, J. M. L. Martin and N. C. Handy "The role of the basis set: Assessing density functional theory" *J. Chem. Phys.* **119**(6), pp. 3005–3014 (2003). doi: **10.1063/1.1589004** (see p. **45**)

[**130**] K. S. Thanthiriwatte, E. G. Hohenstein, L. A. Burns and C. D. Sherrill "Assessment of the Performance of DFT and DFT-D Methods for Describing Distance Dependence of Hydrogen-Bonded Interactions" *J. Chem. Theory Compt.* **7**(1), pp. 88–96 (2011). doi: **10.1021/ct100469b** (see pp. **45**, **50**)

[**131**] J. Ireta, J. Neugebauer and M. Scheffler "On the Accuracy of DFT for Describing Hydrogen Bonds: Dependence on the Bond Directionality" *J. Chem. Phys. A*, **108**(26), pp. 5692–5698 (2004). doi: **10.1021/jp0377073** (see pp. **45**, **50**)

[**132**] C. Kittel, *Introduction to Solid State Physics*. John Wiley & Sons, (2004), p. 704 (see p. **46**)

[**133**] N. W. Ashcroft and N. D. Mermin, *Solid State Physics*. Brooks/Cole, (1976), p. 848 (see pp. **46**, **48**)

[**134**] D. Chadi and Marvin Cohen "Special Points in the Brillouin Zone" *Phys. Rev. B*, **8**(12), pp. 5747–5753 (1973). doi: **10.1103/PhysRevB.8.5747** (see p. **47**)

[**135**] H. J. Monkhorst and J. D. Pack "Special points for Brillouin-zone integrations" *Phys. Rev. B*, **13**(12), pp. 5188–5192 (1976). doi: **10.1103/PhysRevB.13.5188** (see p. **47**)

[**136**] J. R. Chelikowsky "The pseudopotential-density functional method applied to nanostructures" *J. Phys. D. Appl. Phys.* **33**(8), R33–R50 (2000). doi: **10.1088/0022-3727/33/8/201** (see p. **50**)

[**137**] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison "Avogadro: An advanced semantic chemical editor, visualization, and analysis platform" *J. Cheminf.* **4**(1), pp. 1–17 (2012), url: **jcheminf.com/content/4/1/17** (see pp. **54**, **83**)

[**138**] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby and M. Schütz "Molpro: a general-purpose quantum chemistry program package" *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2**(2), pp. 242–253 (2012). doi: **10.1002/wcms.82** (see p. **55**)

[**139**] H. F. Trotter "On the product of semi-groups of operators" *Proc. Am. Math. Soc.* **10**(4), pp. 545–545 (1959). doi: **10.1090/S0002-9939-1959-0108732-6** (see p. **61**)

[**140**] R. W. Hall and B. J. Berne "Nonergodicity in path integral molecular dynamics" *J. Chem. Phys.* **81**(8), p. 3641 (1984). doi: **10.1063/1.448112** (see p. **64**)

[**141**] J. Cao and G. A. Voth "The formulation of quantum statistical mechanics based on the Feynman path centroid density. IV. Algorithms for centroid molecular dynamics" *J. Chem. Phys.* **101**(7), pp. 6168–6183 (1994). doi: **10.1063/1.468399** (see p. **65**)

[**142**] M. E. Tuckerman, B. J. Berne, G. J. Martyna and M. L. Klein "Efficient molecular dynamics and hybrid Monte Carlo algorithms for path integrals" *J. Chem. Phys.* **99**(4), pp. 2796–2808 (1993). doi: **10.1063/1.465188** (see p. **65**)

[**143**] E. Pollock and D. Ceperley "Simulation of quantum many-body systems by path-integral methods" *Phys. Rev. B*, **30**(5), pp. 2555–2568 (1984). doi: **10.1103/PhysRevB.30.2555** (see p. **65**)

[**144**] S. Kathmann, G. Schenter and B. Garrett "Comment on "Quantum Nature of the Sign Preference in Ion-Induced Nucleation"" *Phys. Rev. Lett.* **98**, (2007). doi: **10.1103/PhysRevLett.98.109603** (see p. **72**)

[**145**] S. Kathmann, G. Schenter and B. Garrett "The Critical Role of Anharmonicity in Aqueous Ionic Clusters Relevant to Nucleation" *J. Phys. Chem. C*, **111**(13), pp. 4977–4983 (2007). doi: **10.1021/jp067468u** (see p. **72**)

[**146**] X.-Z. Li, B. Walker and A. Michaelides "Quantum nature of the hydrogen bond" *Proceedings of the National Academy of Sciences*, **108**(16), pp. 6369–6373 (2011). doi: **10.1073/pnas.1016653108** (see pp. **75**, **124**)

[**147**] R. Marcus "Electron transfer reactions in chemistry. Theory and experiment" *Rev. Mod. Phys.* **65**(3), pp. 599–610 (1993). doi: **10.1103/RevModPhys.65.599** (see p. **79**)

[**148**] Y. Wu, H. L. Tepper and G. A. Voth "Flexible simple point-charge water model with improved liquid-state properties." *J. Chem. Phys.* **124**(2), 024503, (2006). doi: **10.1063/1.2136877** (see p. **80**)

[**149**] B. Schiott "On the electronic nature of low-barrier hydrogen bonds in enzymatic reactions" *Proceedings of the National Academy of Sciences*, **95**, pp. 12799–12802 (1998). doi: **10.1073/pnas.95.22.12799** (see p. **81**)

[**150**] M. A. Bellucci and D. F. Coker "Empirical valence bond models for reactive potential energy surfaces: a parallel multilevel genetic program approach." *J. Chem. Phys.* **135**(4), p. 044115 (2011). doi: **10.1063/1.3610907** (see p. **81**)

[**151**] F. Wang and G. A. Voth "A linear-scaling self-consistent generalization of the multistate empirical valence bond method for multiple excess protons in aqueous systems" *J. Chem. Phys.* **122**(14), 144105, (2005). doi: **10.1063/1.1881092** (see pp. **82**, **88**, **90**)

[**152**] V. Loukonen, T. Kurtén, I. K. Ortega, H. Vehkamäki, A. A. H. Pádua, K. Sellegri and M. Kulmala "Enhancing effect of dimethylamine in sulfuric acid nucleation in the presence of water - a computational study" *Atmos. Chem. Phys.* **10**(10), pp. 4961–4974 (2010). doi: **10.5194/acp-10-4961-2010** (see p. **86**)

[**153**] J. López-Lemus, G. A. Chapela and J. Alejandre "Effect of flexibility on surface tension and coexisting densities of water." *J. Chem. Phys.* **128**(17), p. 174703 (2008). doi: **10.1063/1.2907845** (see p. **86**)

[**154**] R. Poli "Analysis of the Publications on the Applications of Particle Swarm Optimisation" *J. Artif. Evol. App.* **2008**, 685175, (2008). doi: **10.1155/2008/685175** (see pp. **94**, **95**)

[**155**] J. Kennedy and R. Eberhart. "Particle swarm optimization". *Proceedings of ICNN'95 - International Conference on Neural Networks*. Vol. 4. IEEE, (1995), pp. 1942–1948. doi: **10.1109/ICNN.1995.488968** (see pp. **94**, **95**)

[**156**] Y. Shi and R. Eberhart. "A modified particle swarm optimizer". *Proceedings of the IEEE International Conference on Evolutionary Computation*. IEEE, (1998), pp. 69–73. doi: **10.1109/ICEC.1998.699146** (see pp. **94**, **96**)

[**157**] C. M. Fonseca and P. J. Fleming "An Overview of Evolutionary Algorithms in Multiobjective Optimization" *Evol. Comput.* **3**(1), pp. 1–16 (1995). doi: **10.1162/evco.1995.3.1.1** (see p. **94**)

[**158**] A. Banks, J. Vincent and C. Anyakoha "A review of particle swarm optimization. Part I: background and development" *Nat. Comput.* **6**(4), pp. 467–484 (2007). doi: **10.1007/s11047-007-9049-5** (see pp. **94**, **99**, **105**)

[**159**] A. Banks, J. Vincent and C. Anyakoha "A review of particle swarm optimization. Part II: hybridisation, combinatorial, multicriteria and constrained optimization, and indicative applications" *Nat. Comput.* **7**(1), pp. 109–124 (2007). doi: **10.1007/s11047-007-9050-z** (see pp. **94**, **99**, **105**)

[**160**] D. A. McQuarrie, *Mathematical Methods for Scientists and Engineers*. University Science Books, (2003), p. 1184 (see p. **144**)

Appendix A

Canonical temperature distribution

The question of the gaining the correct temperature distribution from a MD simulation is addressed in this appendix.

A.1 Variance of system temperature

The Maxwell-Boltzmann distribution [**109**] describes the expected distribution of speeds observed within a system of perfect gas at equilibrium. It is the expected distribution for the Langevin thermostats in use with simulations of NVT ensembles. This distribution has the following form

$$f(v)dv = \sqrt{\left(\frac{m}{2\pi k_B T}\right)^3} 4\pi v^2 \exp\left(\frac{-mv^2}{2k_B T}\right) dv \tag{A.1.1}$$

It is possible to derive the mean value of any property of the system which is dependent on this distribution such as the mean temperature, $\langle T \rangle$ where

$$\begin{aligned} \frac{3k_B T}{2} &= \frac{1}{2}mv^2 \\ T &= \frac{m}{3k_B}v^2 \end{aligned}$$

The derivation will start with one particle in a three dimensional space. To find the mean for the temperature we must first multiply it by the distribution function, Eq. (**A.1.1**) and then integrate over

all possible values such that,

$$\langle T \rangle = \int_0^\infty \frac{m}{3k_B} v^2 \sqrt{\left(\frac{m}{2\pi k_B T_R}\right)^3} 4\pi v^2 \exp\left(\frac{-mv^2}{2k_B T_R}\right) dv \tag{A.1.2}$$

where T_R is the temperature of the reservoir, this reservoir temperature is the target temperature for the thermostat. Eq. (**A.1.2**) can be evaluated using the standard integral [**160**],

$$\int_0^\infty x^{2n} \exp\left(-ax^2\right) dx = \frac{1 \cdot 3 \cdot 5 \cdot (2n-1)}{2^{n+1} a^n} \left(\frac{\pi}{a}\right)^{1/2}$$

We find the result to be as expected,

$$\langle T \rangle = T_R$$

We are now interested in the expected distribution of instantaneous temperatures within a NVT ensemble, i.e. the variance of temperatures, $\left\langle (T - \langle T \rangle)^2 \right\rangle$, this can be evaluated

$$(T - \langle T \rangle)^2 = \frac{1}{N} \sum_{i=1}^{N} (T_i - \langle T \rangle) = \left(\frac{m}{3k_B} v^2 - T_R\right)^2$$

Again we use Eq. (**A.1.1**) to allow us to write,

$$\left\langle (T - \langle T \rangle)^2 \right\rangle = \int_0^\infty \left(\frac{m}{3k_B} v^2 - T_R\right)^2 \sqrt{\left(\frac{m}{2\pi k_B T_R}\right)^3} 4\pi v^2 \exp\left(\frac{-mv^2}{2k_B T_R}\right) dv \tag{A.1.3}$$

This integral can be evaluated as before to give,

$$\left\langle (T - \langle T \rangle)^2 \right\rangle = \frac{2T_R^2}{3} \tag{A.1.4}$$

This variance is related to a Gaussian temperature distribution of the general form

$$\mathcal{P}(T) dT = \sqrt{\frac{1}{\sigma_T 2\pi}} \exp\left(-\frac{1}{2\sigma_T} (T - T_R)^2\right) dT \tag{A.1.5}$$

where $\mathcal{P}(T)$ is the probability distribution of the system temperature. We are interested in the σ_T is the standard deviation and is given by

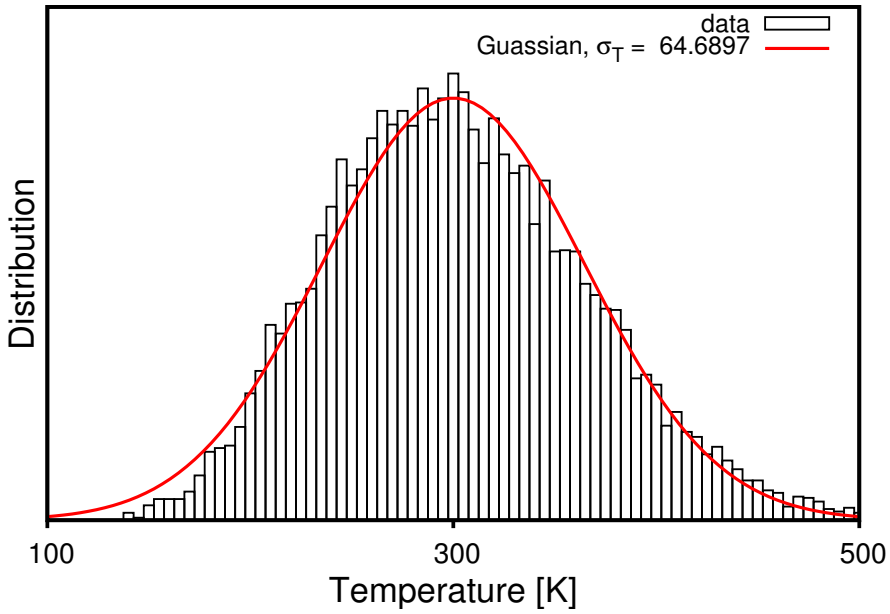


Figure A.1: Comparison of temperature histogram of a DFT MD simulation run at 300 K against Gaussian where the standard deviated was fitted to the value of $\sigma = 64.6897$ K using the gnuplot program (version 4.4), the simulation has a total of 19 atoms and so standard deviation is expected to be of order $\sigma_{T,sys} = 300 \times (2/(3 \times 19))^{1/2} = 56.1951$ K.

$$\frac{\sigma_T}{T_R} = \left(\frac{2}{3}\right)^{1/2}$$

At this stage we turn our attention back to a system made up of N particles. We assume that the speed of each particle is uncorrelated and because each particle has the same variance we can write [**108**, **109**],

$$\begin{aligned} \sigma^2_{T,sys} &= \frac{1}{N}\sigma^2_T \\ \frac{\sigma_{T,sys}}{T_R} &= \left(\frac{2}{3N}\right)^{1/2} \end{aligned} \tag{A.1.6}$$

In practice this is not ideally realised due to limited statistics and imperfections in the application of the chosen thermostat, however Eq. (**A.1.6**) can be used as a general test of the thermostat to be assured that the correct temperature distribution is being observed. For instance a histogram from a DFT MD simulation such as one ran in Chapter **4** is studied in Figure **A.1** and found to sufficiently satisfy Eq. (**A.1.6**).

A.2 The equipartition theorem

Given the issues experienced in Section **3.5** it was decided that further simulations of sulphuric acid and water clusters would be run with no constraint imposed upon its centre of mass. To check that the simulations were now running as expected the kinetic energy was examined to see if the equipartition theorem was being obeyed. The equipartition theorem states that the kinetic energy of a cluster gives rise to three degrees of freedom each for the rotation of the cluster and the translation of the centre of mass for the cluster. The remaining $3N-6$ degrees of freedom to be contributed from internal degrees of freedom. The kinetic energy of the centre of mass translation is trivial to calculate such that,

$$p_{cluster} = \sum_{i=1}^{N} m_i v_i$$

$$T_{COM} = \frac{p_{cluster}^2}{2m_{cluster}} \tag{A.2.1}$$

where $p_{cluster}$ and $m_{cluster}$ is the net momentum of the cluster and the mass of the cluster respectively. T_{COM} is the kinetic energy of the centre of mass of the cluster. i is a label for each atom and there is a total of N atoms in the cluster. To calculate the rotational kinetic energy of the cluster it is useful to start with calculating the angular moment of the cluster, $\mathbf{L}$,

$$\mathbf{L} = \sum_{i=1}^{N} \mathbf{r'}_i \times (m_i \mathbf{v}_i) \tag{A.2.2}$$

where $\mathbf{r'}_i = \mathbf{r}_i - \mathbf{r}_{COM}$ and $\mathbf{r}_{COM}$ is the position vector of the centre of mass for the cluster. Next the moment of inertia matrix for the cluster can be computed as,

$$\underline{\underline{I}} = \begin{bmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{xy} & I_{yy} & I_{yz} \\ I_{xz} & I_{yx} & I_{zz} \end{bmatrix} \tag{A.2.3}$$

where the off diagonals of $\underline{\underline{I}}$ are defined by $I_{ab} = -\sum_{i=1}^{N} m_i \left(\mathbf{r}'_{i,a} \mathbf{r}'_{i,b} \right)$ such that a and b represent the cartesian coordinates x, y and z of vector $\mathbf{r}'_i$ and $a \neq b$. The diagonal components of $\underline{\underline{I}}$ are defined by $I_{aa} = \sum_{i=1}^{N} m_i \left(\left(\mathbf{r}'_{i,b} \right)^2 + \left(\mathbf{r}'_{i,c} \right)^2 \right)$ where $a \neq b$, $a \neq c$ and $b \neq c$. The angular momentum of the

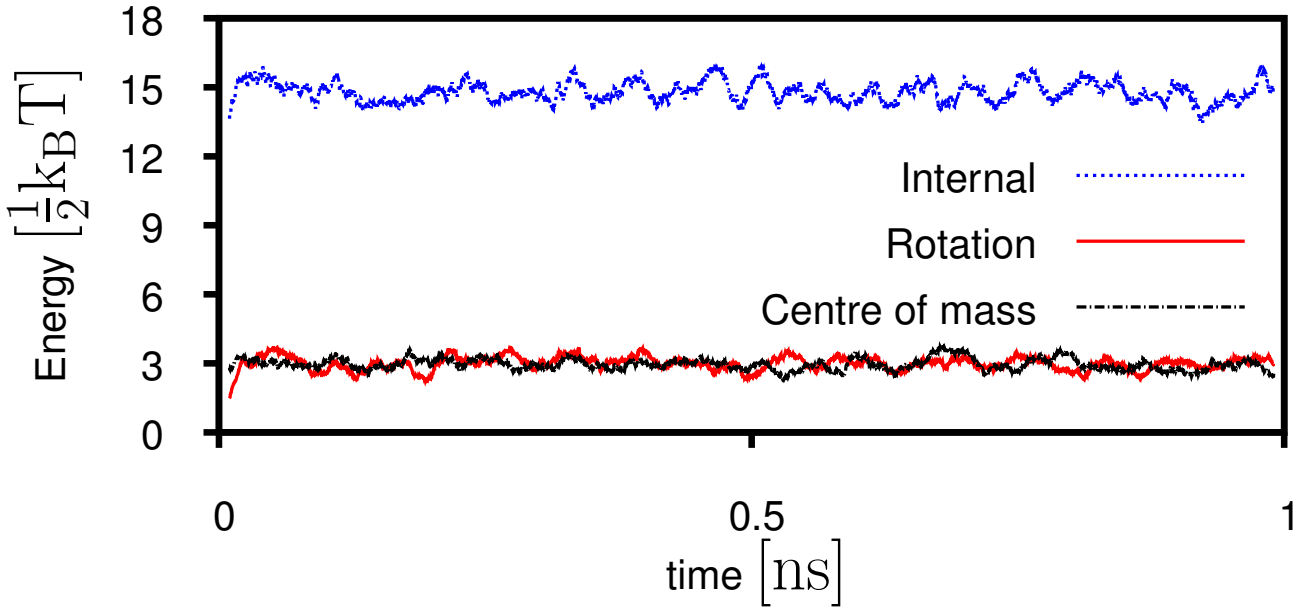


Figure A.2: Image shows the deconstruction of the kinetic energy of a sulphuric acid molecules into three separate energy contributions. These are namely the rotation, translation and internal kinetic energy, plotted as a running average over 100 steps. The plot follows that which is predicted from the equipartition theorem, i.e. rotation and translation account for 3 degree of freedom each and the internal kinetic energy accounts for the remaining 15 degrees of freedom. Each of these degrees of freedom contribute a $\frac{1}{2}k_B T$ to the kinetic energy as seen here.

cluster and the moment of inertia of the cluster are related by the following expression,

$$\mathbf{L} = \underline{\underline{I}}\omega \tag{A.2.4}$$

where ω is the angular velocity of the cluster. To calculate this we multiply Eq. (**A.2.4**) by the inverse of the moment of inertia matrix to get,

$$\omega = \underline{\underline{I}}^{-1}\mathbf{L} \tag{A.2.5}$$

We can now calculate the rotational kinetic energy of the cluster, T_{ROT}, in the following way,

$$T_{ROT} = \frac{1}{2}\omega^{\intercal}\underline{\underline{I}}\omega \tag{A.2.6}$$

Finally the internal kinetic energy of the cluster is simply the total kinetic energy minus T_{COM} and T_{ROT} such that,

$$T_{internal} = \left(\sum_{i=1}^{N}\frac{1}{2}m_i v_i^2\right) - T_{COM} - T_{ROT} \tag{A.2.7}$$

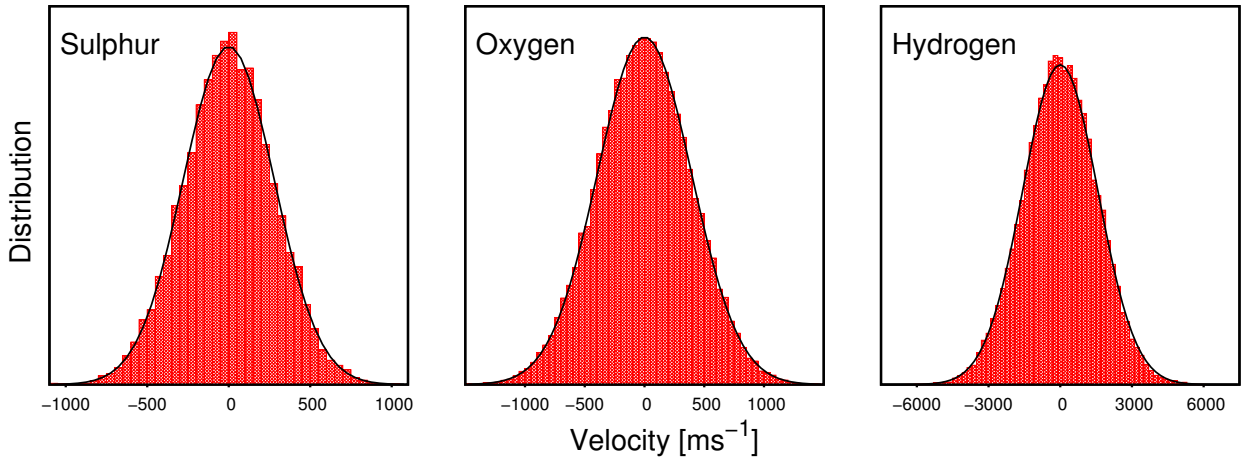


Figure A.3: Normalised histogram of the velocities of various atoms from the same simulation that produced Figure **A.2**. The black line indicates the expected shape of the histogram according to the Maxwell-Boltzmann distribution as described by Eq. (**A.3.2**). The histogram and the theory match up well indicating that the thermostat is performing well.

To test that the Langevin thermostat was correctly obeying the equipartition theorem in the simulations where the centre of mass was constrained a single sulphuric acid molecule was simulated for 1 ns and the three separate contributions to the kinetic energy of the system was monitored. The results are shown in Figure **A.2** and are plotted in units of $\frac{1}{2}k_BT$, it is clear that this simulation obeys the equipartition theorem.

A.3 Species temperature distributions

One final check that the thermostat is performing as expected is to test that it is producing the correct distribution of velocities for the constitute atoms. The correct distribution for that of an NVT ensemble was detailed in Section **A.1**, the expected single component velocity distribution is expected to take the form of [**109**],

$$a = b \tag{A.3.1}$$

$$f(v_x) = \sqrt{\frac{m}{2\pi k_B T}} \exp\left(-\frac{mv_x^2}{2k_B T}\right) \tag{A.3.2}$$

where v_x describes the x-component of the velocity, this holds also for the y and z components. The simulations performed on DL_POLY 4.03 with the removed constraint upon the centre of mass motion do indeed follow this distribution for the velocity of the composite atoms. An example from

the same simulation that produced Figure **A.2** is shown in Figure **A.3** for each of the atoms involved in the simulation.

Appendix B

The Hohenberg-Kohn theorems

In 1964 Hohenberg and Kohn [**51**] introduced the modern version of electronic density functional theory with the Hohenberg-Kohn theorems. Proofs for these two founding theorems (as described in Section **4.1.1**) are presented here.

B.1 Theorem I

Theorem I states that the external potential $V_{ext}(\mathbf{r})$ is uniquely determinable, up to a constant, from the ground state particle density $n_0(\mathbf{r})$. A simple approach to demonstrate this theorem is to start with two separate external potentials $V^1_{ext}(\mathbf{r})$ and $V^2_{ext}(\mathbf{r})$ which differ by more than a constant. These potential are related to separate Hamiltonians, $\mathcal{H}^n,\ n = 1, 2$ and separate ground state wavefunction, Ψ^n. It is then possible to state the following inequality,

$$E^1 = \langle \Psi^1 | \, \mathcal{H}^1 \, | \Psi^1 \rangle < \langle \Psi^2 | \, \mathcal{H}^1 \, | \Psi^2 \rangle \tag{B.1.1}$$

where E^n represents the ground state energy of system n. Assuming both wavefunctions give the same ground state density, $n_0(\mathbf{r})$, we can then rewrite the right hand side of Eq. (**B.1.1**) such that,

$$\langle \Psi^2 | \, \mathcal{H}^1 \, | \Psi^2 \rangle = \langle \Psi^2 | \, \mathcal{H}^2 \, | \Psi^2 \rangle + \langle \Psi^2 | \, \mathcal{H}^1 - \mathcal{H}^2 \, | \Psi^2 \rangle = E^2 + \int d\mathbf{r} \left[V^1_{ext}(\mathbf{r}) - V^2_{ext}(\mathbf{r}) \right] n_0(\mathbf{r}) \tag{B.1.2}$$

We then use the freedom to swap labels ($1 \leftrightarrow 2$) to find at the converse inequality. Finally we sum the two inequalities and arrive at the following expression,

$$E^1 + E^2 < E^2 + E^1$$

which is incorrect. Therefore if two external potentials differ by more than a constant then they cannot share the same ground state particle density, which is a just an alternative statement for theorem I of the Hohenberg-Kohn theorems.

B.2 Theorem II

The second theorem states that a *universal functional* of the particle density which is valid for any external potential can be defined such that the particle density that minimises this functional is the ground state density. The proof of this theorem requires first that we specify that the densities of interest are the ground state density for the specific external potential and internal Hamiltonian. In order for the functional to be considered universal it must only be dependent on the particle density. We first specify the total energy functional such that,

$$E_{HK}[n] = T[n] + E_{int}[n] + \int d\mathbf{r} V_{ext}(\mathbf{r}) n(\mathbf{r}) + E_{II}$$

where $T[n]$ is the kinetic energy functional, $E_{int}[n]$ is the electron internal energy and E_{II} is the potential energy arising from nuclei-nuclei interactions. It is useful to group the first two terms, i.e. $F_{HK}[n] = T[n] + E_{int}[n]$. If the form of the functional F_{HK} is known, then minimising E_{HK} with respect to the density would return the exact ground state density thus proving theorem II. One must be careful in practice as the theorem only concern the ground state of the system, in other words, special consideration is required when studying excited states using DFT.